

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:02 ; Search time 177 Seconds
(without alignments)
2638.788 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFVNQFNKVPNGVDIA.....EFIPVDGNGERPLHHHHH 1302

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_23Sep04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6756	99.0	1296	2	AAR95010 C. botuli
2	6727	98.6	1295	5	AAU99339 Clostridi
3	4515	66.2	871	2	AAW56008 Botulinum
4	4515	66.2	871	8	ADL90290 Clostridi
5	4512	66.1	871	8	ADL90437 Clostridi
6	4512	66.1	873	8	ADL90447 Clostridi
7	4512	66.1	873	8	ADL90445 Clostridi
8	4509	66.1	873	8	ADL90451 Clostridi
9	4507	66.1	871	2	AAW56019 Recombina
10	4507	66.1	871	2	AAW56007 Recombina
11	4507	66.1	871	8	ADL90435 Clostridi
12	4507	66.1	871	8	ADL90308 Clostridi
13	4507	66.1	871	8	ADL90284 Clostridi
14	4507	66.1	873	8	ADL90433 Clostridi
15	4507	66.1	873	8	ADL90431 Clostridi
16	4507	66.1	879	8	ADL90441 Clostridi
17	4507	66.1	1127	8	ADL90322 Clostridi
18	4507	66.1	1129	8	ADL90324 Clostridi
19	4504	66.0	873	8	ADL90449 Clostridi
20	4501	66.0	1013	2	AAW56013 Recombina
21	4501	66.0	1013	8	ADL90300 Clostridi
22	4495.5	65.9	1130	8	ADL90326 Clostridi
23	4495.5	65.9	1130	8	ADL90421 Clostridi
24	4495.5	65.9	1132	8	ADL90328 Clostridi
25	4495	65.9	873	2	AAW56016 Recombina

ALIGNMENTS

RESULT 1
AAR95010
ID AAR95010 standard; protein; 1296 AA.

XX AAR95010;

AC AAR95010;

XX 09-JUL-1996 (first entry)

XX C. botulinum type A neurotoxin.

XX Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.

XX Clostridium botulinum.

XX WO9612802-A1.

XX 02-MAY-1996.

XX 23-OCT-1995; 95WO-US013737.

XX 24-OCT-1994; 94US-00329154.

XX 16-MAR-1995; 95US-00405496.

XX 14-APR-1995; 95US-00422711.

XX 07-JUN-1995; 95US-00480604.

XX (OPHI-) OPHIDIAN PHARM INC.

XX Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;

XX Firca JR;

XX WPI; 1996-230603/23.

XX N-PSDB; AAT29244.

XX Fusion proteins comprising non-toxin protein and part of toxin - useful

XX to form anti-toxins against Clostridium botulinum type A, and C.

XX difficile type toxins, and to treat C. difficile intoxication, partic.

XX diarrhoea.

XX Claim 4; Page 344-350; 434pp; English.

XX Clostridium botulinum type A neurotoxin (AAR95010) is processed to form a

XX dimer composed of a light and a heavy chain. It is the product of the

XX type A neurotoxin gene (AAT29244). The 50 kDa C-terminal portion of the

XX heavy chain, or C fragment (see also AAR95008), was produced using a

XX synthetic gene (AAT29245) having codon usage altered to improve

XX expression in Escherichia coli. Fusion proteins of the type A toxin or C

XX fragment, with e.g. maltose binding protein or polyhistidine affinity tag

XX (see also AAR95008), are used to generate neutralising antitoxins and in

26	4495	65.9	873	8	ADL90288	Adl90288 C. botuli
27	4495	65.9	875	2	AAW56009	AAW56009 Recombina
28	4495	65.9	875	8	ADL90292	Adl90292 Clostridi
29	4495	65.9	877	8	ADL90439	Adl90439 Clostridi
30	4495	65.9	887	8	ADL90443	Adl90443 Clostridi
31	4495	65.9	894	2	AAW56015	AAW56015 Recombina
32	4495	65.9	894	2	ADL90286	Adl90286 Clostridi
33	4494.5	65.9	907	2	AAW56012	AAW56012 Recombina
34	4494.5	65.9	907	8	ADL90298	Adl90298 Clostridi
35	4494.5	65.9	953	2	AAW56011	AAW56011 Recombina
36	4494.5	65.9	953	8	ADL90296	Adl90296 Clostridi
37	4493.5	65.9	878	2	AAW56010	AAW56010 Recombina
38	4493.5	65.9	878	8	ADL90294	Adl90294 Clostridi
39	4485	65.8	1127	8	ADL90340	Adl90340 Clostridi
40	4485	65.8	1129	8	ADL90338	Adl90338 Clostridi
41	4484.5	65.7	908	8	ADL90346	Adl90346 Clostridi
42	4479	65.7	1127	8	ADL90332	Adl90332 Clostridi
43	4479	65.7	1129	8	ADL90330	Adl90330 Clostridi
44	4473.5	65.6	949	8	ADL90350	Adl90350 Clostridi
45	4473	65.6	914	8	ADL90342	Adl90342 Clostridi

XX	27-NOV-2001; 2001WO-US045059.	QY	482	TSDTNEAABENISLDLIQQYLYTFNPDNEPENISNIENSSDIIGOLELMPNIEFPNGK	541
XX		DB	481	TSDTNEAABENISLDLIQQYLYTFNPDNEPENISNIENSSDIIGOLELMPNIEFPNGK	540
XX	29-NOV-2000; 2000US-00726949.	QY	542	KYELDKYTFPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYKYNKATEAA	601
XX	(ALLR) ALLERGAN SALES INC.	DB	541	KYELDKYTFPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYKYNKATEAA	600
XX	Lin W, Aoki KR, Steward LE;	QY	602	MFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGALNIGNMLYKODFVGALIFSGA	661
XX	WPI; 2002-557531/59.	DB	601	MFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGALNIGNMLYKODFVGALIFSGA	660
XX	Novel modified botulinum toxin or tetanus toxin comprising a protease	QY	662	VILLEFIPEIAIPVLGTFALVSYIANKVLTVQTDNALSQRNEKWEVYKIVITNWLAKV	721
XX	cleavage site, is useful for treating conditions benefited by neurotoxin	DB	661	VILLEFIPEIAIPVLGTFALVSYIANKVLTVQTDNALSQRNEKWEVYKIVITNWLAKV	720
XX	activity.	QY	722	NTQIDILIRKQKALENQAEATKAIINYQYNOVTESEKNNININIDLSKLNESINKAM	781
XX	Disclosure; Fig 1; 19pp; English.	DB	721	NTQIDILIRKQKALENQAEATKAIINYQYNOVTESEKNNININIDLSKLNESINKAM	780
XX	The invention discloses modified botulinum toxin (BoToX) or tetanus toxin	QY	782	ININKELNOCVSYLNMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRKDKV	841
XX	(TeToX) which contain new protease cleavage sites. These sites are	DB	781	ININKELNOCVSYLNMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRKDKV	840
XX	created in regions where the cleavage site is specifically susceptible to	QY	842	NNTLSTDIPQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG	901
XX	cleavage by a protease that is present in effective levels only in a	DB	841	NNTLSTDIPQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG	900
XX	tissue where toxicity is undesirable and where the cleavage site itself	QY	902	SKYVDFDIDKNOIQLFNLESSKIEVLKNAIVNYSMYENFSTFWIRIPKYPNSISLANE	961
XX	does not inactivate the toxin. BoToX and TeToX are neurotoxins which are	DB	901	SKYVDFDIDKNOIQLFNLESSKIEVLKNAIVNYSMYENFSTFWIRIPKYPNSISLANE	960
XX	bond. The modified toxins are useful to treat conditions benefited by	QY	962	YTIINCWNNNSGKVSINLGEIITWLTQDTEIKORVVFKYSOMINISDIYINRWFVTITN	1021
XX	neurotoxin activity (e.g. spastic conditions, including stabismus,	DB	961	YTIINCWNNNSGKVSINLGEIITWLTQDTEIKORVVFKYSOMINISDIYINRWFVTITN	1020
XX	cephalospasm and hemifacial spasm, brain injury, spinal cord injury,	QY	1022	NRLNNSKIYINGRLIDQKPIISNLGNITHASNNIMPKLDGCRDTHRYIWKYFNLFDKELNE	1081
XX	stroke, multiple sclerosis and cerebral palsy) by administering the	DB	1021	NRLNNSKIYINGRLIDQKPIISNLGNITHASNNIMPKLDGCRDTHRYIWKYFNLFDKELNE	1080
XX	toxins for their localised production. The advantage of the toxins are	QY	1082	KEIKDLYDNQNSGILKDFWGDYLDVQDKPYMNLNLDPNKYVDVNNVGIHGYMLKGPGRG	1141
XX	that they are deactivated in tissues where toxic activity is undesirable	DB	1081	KEIKDLYDNQNSGILKDFWGDYLDVQDKPYMNLNLDPNKYVDVNNVGIHGYMLKGPGRG	1140
XX	and activated at desired targets. The sequence presented is the	QY	1142	SVMTTIYLNLSLYRGTKFTIIKKYASGNKDNIVRNNDRVYINVVVKNKYRLATNASQAG	1201
XX	Clostridium botulinum botulinum A toxin (BoNT/A) protein	DB	1141	SVMTTIYLNLSLYRGTKFTIIKKYASGNKDNIVRNNDRVYINVVVKNKYRLATNASQAG	1200
XX	Sequence 1295 AA;	QY	1202	VEKILSALEIPDVGNLSQVVMKSKNDQGITNCKQNLQDNGNDIGFIGHQFNNTAKL	1261
XX	Query Match 98.6%; Score 6727; DB 5; Length 1295;	DB	1201	VEKILSALEIPDVGNLSQVVMKSKNDQGITNCKQNLQDNGNDIGFIGHQFNNTAKL	1260
XX	Best Local Similarity 99.5%; Pred. No. 0;	QY	1262	VASNWRNQRISRTILGCSWEFIPVDDGHWGERPL 1296	
XX	Matches 1288; Conservative 1; Mismatches 6; Indels. 0; Gaps 0;	DB	1261	VASNWRNQRISRTILGCSWEFIPVDDGHWGERPL 1295	
QY	2 PFVNKQFNYPDPVNGVDIAIKIPINAGOMQPVKAFKIHNKIWIWIPERDFTTNBEGDLNP	61			
DB	1 PFVNKQFNYPDPVNGVDIAIKIPINAGOMQPVKAFKIHNKIWIWIPERDFTTNBEGDLNP	60			
QY	62 PPEAKQVPVSYDSTYLSLTDNEKDNLYLKVYTKLFRYISTDLGRMLTSLVRGIPFWGGS	121			
DB	61 PPEAKQVPVSYDSTYLSLTDNEKDNLYLKVYTKLFRYISTDLGRMLTSLVRGIPFWGGS	120			
QY	122 TIDTELKVIDTNCINVTQDGSYRSEELNLVIIGPSADIIQFECKSPGHEVLNLRNGYG	181			
DB	121 TIDTELKVIDTNCINVTQDGSYRSEELNLVIIGPSADIIQFECKSPGHEVLNLRNGYG	180			
QY	182 STQYIRFSPDFTFGFEESLEVDNPLLGAGKFAFDPAVTLAHELIIYAGHRLYIAINPNR	241			
DB	181 STQYIRFSPDFTFGFEESLEVDNPLLGAGKFAFDPAVTLAHELIIYAGHRLYIAINPNR	240			
QY	242 VFKNVTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKPDIASITLNKAK	301			
DB	241 VFKNVTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKPDIASITLNKAK	300			
QY	302 SIVGTTASLQVMKNVFEKYLSDTSKFSVDKLKFDKLYKMLTEIYTEDNFVKFPKVL	361			
DB	301 SIVGTTASLQVMKNVFEKYLSDTSKFSVDKLKFDKLYKMLTEIYTEDNFVKFPKVL	360			
QY	362 NRKTYLNFDAKAVFKINIVPKVNYTYIDGFNLRLNTLAAFNFGQNTENNMFYKLNKFTG	421			
DB	361 NRKTYLNFDAKAVFKINIVPKVNYTYIDGFNLRLNTLAAFNFGQNTENNMFYKLNKFTG	420			
QY	422 LFEPYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGEI	481			
DB	421 LFEPYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGEI	480			

RESULT 3

AAW56008

ID AAW56008 standard; protein; 871 AA.

XX AAW56008;

XX AC AAW56008;

DT 27-JUL-1998 (first entry)

XX Botulinum neurotoxin type A BoNT/A.

XX Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;

XX detection; tetanus; non-toxic; toxin.

XX Clostridium botulinum.

XX

PN W09807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX WPI; 1998-169168/15.
XX N-PSDB; AAV26280.
XX
XX Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.
XX
XX Disclosure; Page 52-54; 137pp; English.
XX
XX The present sequence represents botulinum neurotoxin type A from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
XX Sequence 871 AA;
SQ
Query Match 66.2%; Score 4515; DB 2; Length 871;
Best Local Similarity 99.9%; Pred. No. 4e-257;
Matches 870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 60
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 60
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 120
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 120
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 180
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 180
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 240
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 240
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 300
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 300
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 360
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 360
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 420
61 1 MPFVNKQNPYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 420

QY 421 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
DB 421 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
QY 481 IISDTNIEAAENISLDLIQOYYLTFNFDNEPENISIEIENISLDLIQOYYLTFNFDNEPENIS 540
DB 481 IISDTNIEAAENISLDLIQOYYLTFNFDNEPENISIEIENISLDLIQOYYLTFNFDNEPENIS 540
QY 541 KKYELDKYTMFHYLRAQEFEGHKSRIALTNVNEALLNFSRVYTFSSDYVKKVNKATEA 600
DB 541 KKYELDKYTMFHYLRAQEFEGHKSRIALTNVNEALLNFSRVYTFSSDYVKKVNKATEA 600
QY 601 AMFLGWVQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGLIFSG 660
DB 601 AMFLGWVQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGLIFSG 660
QY 661 AVLLSEFIPEIATPVLTGTFALSYIANKVLTWTIDNALSKRNEKWDVVKYIVTNWLAK 720
DB 661 AVLLSEFIPEIATPVLTGTFALSYIANKVLTWTIDNALSKRNEKWDVVKYIVTNWLAK 720
QY 721 VNTQIDILIRKMKKEALENOAEATKAIINQYNOYTEEKNINFNIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKMKKEALENOAEATKAIINQYNOYTEEKNINFNIDDLSSKLNESINKA 780
QY 781 MININKPLNOCQSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTILIGQVDRDKDK 840
DB 781 MININKPLNOCQSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTILIGQVDRDKDK 840
QY 841 VNNLTSTDPFQLSKYVDNORLLSTFTYIK 871
DB 841 VNNLTSTDPFQLSKYVDNORLLSTFTYIK 871
RESULT 4
ADL90290
ID ADL90290 standard; protein; 871 AA.
XX
AC ADL90290;
XX
DT 17-JUN-2004 (first entry)
XX
DE Clostridium botulinum BoNT/A protein SEQ ID NO:8.
XX
XX single chain polypeptide; clostridial neurotoxin light chain;
KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
KW botulinum; tetanus.
XX
OS Clostridium botulinum.
XX
PN W02004024909-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-GB003824.
XX
PR 12-SEP-2002; 2002US-00241596.
XX
XX (HEAL-) HEALTH PROTECTION AGENCY.
XX
XX Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
PI Wayne J;
XX
XX WPI; 2004-270039/25.
XX
XX N-PSDB; ADL90289.
XX
XX New single chain polypeptides comprising clostridial neurotoxin light and
PT heavy chains, useful as positive controls for toxin assays, or for
PT developing vaccines against clostridial toxin.
XX
XX Disclosure; SEQ ID NO 8; 588pp; English.
XX

The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

Sequence 871 AA;

Query Match	66.2%; Score 4515; DB 8; Length 871;
Best Local Similarity	99.9%; Pred. No. 4e-257;
Matches	870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MPFVNQFNKDPVNGVDIAIKIPNAGQMPQVAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Db	1 MPFVNQFNKDPVNGVDIAIKIPNAGQMPQVAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Qy	61 PPPEAKQVPVSYDSTYLTSTNEDKNDYIKGVTKLFIERYSTDLGRMLTSLVRGIPFWGG 120
Db	61 PPPEAKQVPVSYDSTYLTSTNEDKNDYIKGVTKLFIERYSTDLGRMLTSLVRGIPFWGG 120
Qy	121 STIDTELKVIDTNCINVIQDGSYRSBELNLVIIGPSADIIQFECKSGFGEVLNLTNGY 180
Db	121 STIDTELKVIDTNCINVIQDGSYRSBELNLVIIGPSADIIQFECKSGFGEVLNLTNGY 180
Qy	181 GSTOYIRFSPDFTGFBESLEVDNTNLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db	181 GSTOYIRFSPDFTGFBESLEVDNTNLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Qy	241 RVFKVNTNAYEMSGLEVSPELTFGHDGDAKFDLSIQENEFRLYYNKKFDIASTLNKA 300
Db	241 RVFKVNTNAYEMSGLEVSPELTFGHDGDAKFDLSIQENEFRLYYNKKFDIASTLNKA 300
Qy	301 KSIIVGTTASLQYMNKVFKEKYLLEDTSFGKPSVDKLFKLYKMLTEIYTDNFKVFPKV 360
Db	301 KSIIVGTTASLQYMNKVFKEKYLLEDTSFGKPSVDKLFKLYKMLTEIYTDNFKVFPKV 360
Qy	361 LNRKTYLNFDFKAVKINIVPKNTIYDGFNRLNLTNLAANFNGQNTNINNMFVKLNFT 420
Db	361 LNRKTYLNFDFKAVKINIVPKNTIYDGFNRLNLTNLAANFNGQNTNINNMFVKLNFT 420
Qy	421 GLFFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPEDNFTDLNKEE 480
Db	421 GLFFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPEDNFTDLNKEE 480
Qy	481 ITSNTNIEAENISLDLIQYIYITFTNFDNENISINENSSDIIGOLELMPNTERPFG 540
Db	481 ITSNTNIEAENISLDLIQYIYITFTNFDNENISINENSSDIIGOLELMPNTERPFG 540
Qy	541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Db	541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Qy	601 AMFLGWELQVYDFTDTSVSTTDKADITIIPIYGPALNIGNMLYKODFVGLIFSG 660
Db	601 AMFLGWELQVYDFTDTSVSTTDKADITIIPIYGPALNIGNMLYKODFVGLIFSG 660

Qy	661 AVILLSEPIPIAIPVLGTFFALVSIANKVLTQVOTIDNALSKRNEKWDVTKYIVTNWLAK 720
Db	661 AVILLSEPIPIAIPVLGTFFALVSIANKVLTQVOTIDNALSKRNEKWDVTKYIVTNWLAK 720
Qy	721 VNTQIDILIRKMKKEALENQAEATKAIINYOYNOVTEEEKNNINFNIDDLSSKLNESINKA 780
Db	721 VNTQIDILIRKMKKEALENQAEATKAIINYOYNOVTEEEKNNINFNIDDLSSKLNESINKA 780
Qy	781 MININKFLNQCYSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTLIGQVDRKLKDK 840
Db	781 MININKFLNQCYSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTLIGQVDRKLKDK 840
Qy	841 VNNLTSTIDIPFQLSKYVDNQRLLSTFTTEYIK 871
Db	841 VNNLTSTIDIPFQLSKYVDNQRLLSTFTTEYIK 871
RESULT 5	
ADL90437	
ID	ADL90437 standard; protein; 871 AA.
XX	AC ADL90437;
XX	17-JUN-2004 (first entry)
DE	Clostridial neurotoxin amino acid sequence SEQ ID NO:155.
KW	single chain polypeptide; clostridial neurotoxin light chain;
KW	Clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW	antibacterial; vaccine; toxin assay; clostridial toxin; detection;
OS	Clostridium botulinum.
PN	WO2004024909-A2.
PD	25-MAR-2004.
PF	12-SEP-2003; 2003WO-GB003824.
PR	12-SEP-2002; 2002US-00241596.
PA	(HEAL-) HEALTH PROTECTION AGENCY.
PI	Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
PI	Wayne J;
XX	WPI; 2004-270039/25.
DR	N-PSDB; ADL90436.
XX	New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for developing vaccines against clostridial toxin.
XX	Claim 1; SEQ ID NO 155; 588pp; English.
CC	The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain

polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

Sequence 871 AA;

Query Match	66.1%	Score 4512	DB 8	Length 871
Best Local Similarity	99.9%	Pred. No. 6.1e-257		
Matches 870	Conservative 0	Mismatches 25	Indels 0	Gaps 0
Qy	1	MPFVKNQFNKYKDPVNGVDIAVIKIPNAGOMQPVKAFKIHNKIWIIPERDTFTNPEEGDLN	60	
Db	1	MEFVKNQFNKYKDPVNGVDIAVIKIPNAGOMQPVKAFKIHNKIWIIPERDTFTNPEEGDLN	60	
Qy	61	PPPEAKQVPVSYYDSTYLDSTNDEKXNYLKGVTYKLFERIYSTDLGRMLITSIVRGIPFWGG	120	
Db	61	PPPEAKQVPVSYYDSTYLDSTNDEKXNYLKGVTYKLFERIYSTDLGRMLITSIVRGIPFWGG	120	
Qy	121	STIDTELKVIDTNCINVIQPDGYSRSEBELNVIIPGSADIIOFECKSFGEHVLNLTNGY	180	
Db	121	STIDTELKVIDTNCINVIQPDGYSRSEBELNVIIPGSADIIOFECKSFGEHVLNLTNGY	180	
Qy	181	GSTQYIRFSDPFTFGFBESLEVDNPLLGAGKFPATDPAVTLAHELIYAGHRLYGIANPN	240	
Db	181	GSTQYIRFSDPFTFGFBESLEVDNPLLGAGKFPATDPAVTLAHELIYAGHRLYGIANPN	240	
Qy	241	RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSJQENEFRLYYNKKFDIASTLNKA	300	
Db	241	RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSJQENEFRLYYNKKFDIASTLNKA	300	
Qy	301	KSIYVGTTSASLOYMKNVFKBYLLSEDTSGKFSVDKLKPDKLYKMLTEIYTEDNFVKFPKV	360	
Db	301	KSIYVGTTSASLOYMKNVFKBYLLSEDTSGKFSVDKLKPDKLYKMLTEIYTEDNFVKFPKV	360	
Qy	361	LNRTYLNFDKAVPKINIVPKVNTIYDGFNLRTNLNLAANFNGQTEINNNWFYTKLNKFT	420	
Db	361	LNRTYLNFDKAVPKINIVPKVNTIYDGFNLRTNLNLAANFNGQTEINNNWFYTKLNKFT	420	
Qy	421	GLFBEFYKLLCVRGIIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSDNFTNDLNKGE	480	
Db	421	GLFBEFYKLLCVRGIIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSDNFTNDLNKGE	480	
Qy	481	ITSDTNIEAABENISLDLIQQYYITFNPDNPNENISIENLSSDIIGOLELMPNTERPPNG	540	
Db	481	ITSDTNIEAABENISLDLIQQYYITFNPDNPNENISIENLSSDIIGOLELMPNTERPPNG	540	
Qy	541	KKYELDKYTMFHYLRAQBFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKVKNKATEA	600	
Db	541	KKYELDKYTMFHYLRAQBFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKVKNKATEA	600	
Qy	601	AMFLGWEOQLVYDFTDETSYSTTDKADIITIIPIYIGPALNIGNMLYKDFVGCALIFSG	660	
Db	601	AMFLGWEOQLVYDFTDETSYSTTDKADIITIIPIYIGPALNIGNMLYKDFVGCALIFSG	660	
Qy	661	AVILLEFIPEIAIPVLGTFALVSYTANKVLTVQTDIDNALSXRNEKWDVEYKIVITNMLAK	720	
Db	661	AVILLEFIPEIAIPVLGTFALVSYTANKVLTVQTDIDNALSXRNEKWDVEYKIVITNMLAK	720	
Qy	721	VNTQIDILRKMKKEALENQAEATKAIINYQNYQVTEEEKNNINFNIDLSKKNESINKA	780	
Db	721	VNTQIDILRKMKKEALENQAEATKAIINYQNYQVTEEEKNNINFNIDLSKKNESINKA	780	
Qy	781	MININKFNLQCSVSYLNMMSIPIYGVKRLDEPDASLKDALLKYIYDNRGTLLIGQVDRLKDK	840	
Db	781	MININKFNLQCSVSYLNMMSIPIYGVKRLDEPDASLKDALLKYIYDNRGTLLIGQVDRLKDK	840	
Qy	841	VNNTLSTDIPFQLSKYVDNQRLLSTFTFYIK	871	
Db	841	VNNTLSTDIPFQLSKYVDNQRLLSTFTFYIK	871	

RESULT 6

ADL90447

ID ADL90447 standard; protein; 873 AA.

XX

AC ADL90447;

XX

DT 17-JUN-2004 (first entry)

XX

DE Clostridial neurotoxin amino acid sequence SEQ ID NO:165.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

KW single chain polypeptide

KW clostridial neur

KW antibacteria

KW botulinum; tetanus. . .
KW
WV

SECRET

OS Clostridium botulinum.
yy

XX
END

PN
YY
W0200402

XX
PD
25-MAR-2004

FD XX
23-MAR-2004.

XX	12-SEP-2003; 2003WO-GB003824.	
PF		
XX		
XX		
PR	12-SEP-2002; 2002US-00241596.	
XX		
XX	(HEAL-) HEALTH PROTECTION AGENCY.	
PA		
XX		
PI	Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;	
PI	Wayne J;	
XX		
XX	WPI: 2004-270039/25.	
DR		
DR	N-PSDB; ADL90445.	
XX		
PT	New single chain polypeptides comprising clostridial neurotoxin light and	
PT	heavy chains, useful as positive controls for toxin assays, or for	
PT	developing vaccines against clostridial toxin.	

PS Claim 1: SEO ID NO 165: 588pp: English:

The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

Sequence 873 AA:

Query Match	66.1%	Score 4512	DB 8	Length 873
-------------	-------	------------	------	------------

Query Match 86.1%; Score 4312; DB 6;
Best Local Similarity 99.9%; Pred. No. 6.1e-257;

Best local similarity 55.5%, pica: no: 0.10 257)
Matches 870: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

1 MPFVNKOFNYKDPVNGVDIA YIKIPNAGOMOPVKAFKIHNKIWV IPERDTFTNPEEGDLN 60

Db 3 MEFVKNQFNKDPVNGVDIAIYKIPNAGOMQPVKAFKIHNKIWIPIPERDITFTNPEEGDLN 62
Qy 61 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
Db 63 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 122
Qy 121 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSGFHEVLNLTNGY 180
Db 123 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSGFHEVLNLTNGY 182
Qy 181 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPVTLAHLIYAGHRLYGIAPNP 240
Db 183 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPVTLAHLIYAGHRLYGIAPNP 242
Qy 241 RVFKVNTNAYYEMSGLEVSFEELRTFGHDADFIDSLQENEFRLYYNKKFDIATLNKA 300
Db 243 RVFKVNTNAYYEMSGLEVSFEELRTFGHDADFIDSLQENEFRLYYNKKFDIATLNKA 302
Qy 301 KSIQVTTASLOYKMNVPKEKYLSEDTSKFSVDKLFKDKLYKMLTEIYTDENFVKPKV 360
Db 303 KSIQVTTASLOYKMNVPKEKYLSEDTSKFSVDKLFKDKLYKMLTEIYTDENFVKPKV 362
Qy 361 LNRKTYLNFDAKFKINIVPKVNTIYDGFNLRTNLAANFNGQNTNINNNTFKLNFT 420
Db 363 LNRKTYLNFDAKFKINIVPKVNTIYDGFNLRTNLAANFNGQNTNINNNTFKLNFT 422
Qy 421 GLFEFYKLLCVRGITSTKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGE 480
Db 423 GLFEFYKLLCVRGITSTKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGE 482
Qy 481 ITSDNTIEAARENISLDLIQOYLYTFNPDNEPENISLENLSSDIIGOLELMPNTERFPNG 540
Db 483 ITSDNTIEAARENISLDLIQOYLYTFNPDNEPENISLENLSSDIIGOLELMPNTERFPNG 542
Qy 541 KKYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Db 543 KKYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 602
Qy 601 AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDFGALLIFSG 660
Db 603 AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDFGALLIFSG 662
Qy 661 AVILLEPTEIPALVGLTFALVSIYANKVLVQIIDNALSKEKWEVYKIIYTNWLAK 720
Db 663 AVILLEPTEIPALVGLTFALVSIYANKVLVQIIDNALSKEKWEVYKIIYTNWLAK 722
Qy 721 VNTQIDLRKKKKEALENOAEATKAIINYQVQYTEBEKNINFNIDDLSSKLNESINKA 780
Db 723 VNTQIDLRKKKKEALENOAEATKAIINYQVQYTEBEKNINFNIDDLSSKLNESINKA 782
Qy 781 MININKFLNQCYSVLYMNSMIPYGVKRLDFDASLKDALKYIYDNRGTLLIGQVDRCLKD 840
Db 783 MININKFLNQCYSVLYMNSMIPYGVKRLDFDASLKDALKYIYDNRGTLLIGQVDRCLKD 842
Qy 841 VNNLTSTDIPQLSKYVDNQRLSTFTFYIK 871
Db 843 VNNLTSTDIPQLSKYVDNQRLSTFTFYIK 873

RESULT 7
ADL90445
ID ADL90445 standard; protein; 873 AA.
XX AC ADL90445;
XX DT 17-JUN-2004 (first entry)

DE Clostridial neurotoxin amino acid sequence SEQ ID NO:163.

XX single chain polypeptide; clostridial neurotoxin light chain;
KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
KW botulinum; tetanus.

XX OS Clostridium botulinum.
XX WO2004024909-A2.
XX PD 25-MAR-2004.
XX PF 12-SEP-2003; 2003WO-GB003824.
XX PR 12-SEP-2002; 2002US-00241596.
XX (HEAL-) HEALTH PROTECTION AGENCY.
XX Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
XX Wayne J;
XX WPI; 2004-270039/25.
XX N-PSDB; ADL90444.
XX New single chain polypeptides comprising clostridial neurotoxin light and
XX heavy chains, useful as positive controls for toxin assays, or for
XX developing vaccines against clostridial toxin.
XX Claim 1; SEQ ID NO 163; 588pp; English.
XX The present invention describes a single chain polypeptide comprising
XX clostridial neurotoxin light and heavy chains. The single chain
XX polypeptide comprises 2 domains: the first domain is a clostridial
XX neurotoxin light chain, or its fragment or variant, which is capable of
XX cleaving one or more vesicle or plasma membrane associated proteins
XX essential to exocytosis; the second domain is a clostridial neurotoxin
XX heavy chain H-N portion, or its fragment or variant, which is capable of
XX translocating the polypeptide into a cell and/or increasing the
XX solubility of the polypeptide compared to the solubility of the first
XX domain on its own. The second domain lacks a functional C-terminal part
XX of a clostridial neurotoxin heavy chain, designated H-C, which renders
XX the polypeptide incapable of binding to cell surface receptors that are
XX the natural cell surface receptors to which native clostridial neurotoxin
XX binds. Also described is a nucleic acid molecule encoding the single
XX chain polypeptide described above. The single chain polypeptide has
XX antibacterial activity, and can be used in vaccines. The single chain
XX polypeptides can be used as positive controls for toxin assays, as
XX reagent components for the synthesis of therapeutic molecules, or for
XX developing vaccines against clostridial toxin. The polypeptides are also
XX useful as non-toxic standards for the assessment and development of in
XX vitro assays for detection of functional botulinum or tetanus neurotoxins
XX in foodstuffs or environmental samples. The present sequence is used in
XX the exemplification of the present invention.
XX SQ Sequence 873 AA;

Query Match 66.1%; Score 4512; DB 8; Length 873;
Best Local Similarity 99.9%; Pred. No. 6.1e-257;
Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MPFVNKQFNKDPVNGVDIAIYKIPNAGOMQPVKAFKIHNKIWIPIPERDITFTNPEEGDLN 60
Db 3 MEFVKNQFNKDPVNGVDIAIYKIPNAGOMQPVKAFKIHNKIWIPIPERDITFTNPEEGDLN 62
Qy 61 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
Db 63 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 122
Qy 121 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSGFHEVLNLTNGY 180
Db 123 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSGFHEVLNLTNGY 182
Qy 181 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPVTLAHLIYAGHRLYGIAPNP 240
Db 183 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPVTLAHLIYAGHRLYGIAPNP 242
Qy 241 RVFKVNTNAYYEMSGLEVSFEELRTFGHDADFIDSLQENEFRLYYNKKFDIATLNKA 300

Db 243 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 302
QY 301 KSIIVGTTASLOYMKNVKPKYLLSEDTSGKFSVDKLFKEDKLYKMLTEIYTDNEDNFKPKV 360
Db 303 KSIIVGTTASLOYMKNVKPKYLLSEDTSGKFSVDKLFKEDKLYKMLTEIYTDNEDNFKPKV 362
QY 361 LNRKTYLNFDAVKFKNIPVKNVYTIYDGFNLRNTNLAANFNGQNTNINNNFTKLNFT 420
Db 363 LNRKTYLNFDAVKFKNIPVKNVYTIYDGFNLRNTNLAANFNGQNTNINNNFTKLNFT 422
QY 421 GLFEFYKLLCVRGIIITSKTSKLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGE 480
Db 423 GLFEFYKLLCVRGIIITSKTSKLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGE 482
QY 481 ITSNTNIEAEBENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNTERP 540
Db 483 ITSNTNIEAEBENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNTERP 542
QY 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Db 543 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 602
QY 601 AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
Db 603 AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 662
QY 661 AVILLEPETAIPVLGTFALVSYANKVLTVQTDNALSXRNKQWDEVYKIVTNWLAK 720
Db 663 AVILLEPETAIPVLGTFALVSYANKVLTVQTDNALSXRNKQWDEVYKIVTNWLAK 722
QY 721 VNTQIDILIRKQWKEALENQAEATKAIINYQNYTEEEKNNINFNIDDLSSKLNESINKA 780
Db 723 VNTQIDILIRKQWKEALENQAEATKAIINYQNYTEEEKNNINFNIDDLSSKLNESINKA 782
QY 781 MININKFLNQCYSVYLMNSMIPYGVKRLIEDPDASIKDALLKYIYDNRGTLIGQVDRLLKDK 840
Db 783 MININKFLNQCYSVYLMNSMIPYGVKRLIEDPDASIKDALLKYIYDNRGTLIGQVDRLLKDK 842
QY 841 VNTLSTDDIPQLSKYVDNQRLISTFTTEYIK 871
Db 843 VNTLSTDDIPQLSKYVDNQRLISTFTTEYIK 873

RESULT 8

ADL90451
ID ADL90451 standard; protein; 873 AA.
XX
AC ADL90451;
XX
DT 17-JUN-2004 (first entry)
XX
DE Clostridial neurotoxin amino acid sequence SEQ ID NO:169.
XX
KW single chain polypeptide; clostridial neurotoxin light chain;
KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
KW botulinum; tetanus.
XX
OS Clostridium botulinum.
XX
FN WO2004024909-A2.
XX
PD 25-MAR-2004.
XX
FF 12-SEP-2003; 2003WO-GB003824.
XX
PR 12-SEP-2002; 2002US-00241596.
XX
PA (HEAL-) HEALTH PROTECTION AGENCY.
XX
FI Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
PI Wayne J;
XX

DR WPI; 2004-270039/25.
XX N-PSDB; ADL90450.

PT New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for developing vaccines against clostridial toxin.

PS Claim 1; SEQ ID NO 169; 588pp; English.

XX The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has chain polypeptide activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

SQ Sequence 873 AA;

Query Match 66.1%; Score 4509; DB 8; Length 873;

Best Local Similarity 99.8%; Pred. No. 9.1e-257;

Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPFVNFQNFYKDPVNGVDIAIYIKIPNAGQMPKAFKIHKNKIWIIPERDFTTPEEGDLN 60

Db 3 MEFVNFQNFYKDPVNGVDIAIYIKIPNAGQMPKAFKIHKNKIWIIPERDFTTPEEGDLN 62

QY 61 PPEAKQVPVSYDYSTYLSLTDNEKNYLGKVTYKLFERIYSTDLGRMLLTISVIRGIPFWGG 120

Db 63 PPEAKQVPVSYDYSTYLSLTDNEKNYLGKVTYKLFERIYSTDLGRMLLTISVIRGIPFWGG 122

QY 121 STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIIGPSADIIQFECKSFGEHVLNLTNGY 180

Db 123 STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIIGPSADIIQFECKSFGEHVLNLTNGY 182

QY 181 GSTQYIRFSPDFTGPFESLEVDTPNLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240

Db 183 GSTQYIRFSPDFTGPFESLEVDTPNLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 242

QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300

Db 243 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 302

QY 301 KSIIVGTTASLOYMKNVKPKYLLSEDTSGKFSVDKLFKEDKLYKMLTEIYTDNEDNFKPKV 360

Db 303 KSIIVGTTASLOYMKNVKPKYLLSEDTSGKFSVDKLFKEDKLYKMLTEIYTDNEDNFKPKV 362

QY 361 LNRKTYLNFDAVKFKNIPVKNVYTIYDGFNLRNTNLAANFNGQNTNINNNFTKLNFT 420

Db 363 LNRKTYLNFDAVKFKNIPVKNVYTIYDGFNLRNTNLAANFNGQNTNINNNFTKLNFT 422

QY 421 GLFEFYKLLCVRGIIITSKTSKLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGE 480

Db 423 GLFEFYKLLCVRGIIITSKTSKLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGE 482

QY 481 ITSNTNIEAEBENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNTERP 540

XX

Db 483 ITSDTNIIEAENISLDLIQQYVLTFTFDPNEPENISLSDIIGOLELMPNTERFPNG 542
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 600
Db 543 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 602
Qy 601 AMELGWVQVYVDFDTSEVSTTDKIADITIIPIYGPALNIGMMLYKDDFVGCALIFSG 660
Db 603 AMELGWVQVYVDFDTSEVSTTDKIADITIIPIYGPALNIGMMLYKDDFVGCALIFSG 662
Qy 661 AVILLEPIEPIAIPVLGTGFALVSIAKVKLTQTIDNALSKEKWKDEVKYIIVTNWLAK 720
Db 663 AVILLEPIEPIAIPVLGTGFALVSIAKVKLTQTIDNALSKEKWKDEVKYIIVTNWLAK 722
Qy 721 VNTQIDILIRKKMEALENQAATKAIINYQVQYTEEEKNNINFNIDLSKKNESINKA 780
Db 723 VNTQIDILIRKKMEALENQAATKAIINYQVQYTEEEKNNINFNIDLSKKNESINKA 782
Qy 781 MININKFLNOCVSYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGQVDRDKDK 840
Db 783 MININKFLNOCVSYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGQVDRDKDK 842
Qy 841 VVNTLSTDIPFQLSKYVDNORLLSTFTTEYIK 871
Db 843 VVNTLSTDIPFQLSKYVDNORLLSTFTTEYIK 873

RESULT 9
AAW56019
ID AAW56019 standard; protein; 871 AA.
AC AAW56019;
XX 27-JUL-1998 (first entry)
XX Recombinant botulinum neurotoxin type A LH423/A (Q2E,N26K,A27Y).
XX Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
XX detection; tetanus; non-toxic; toxin.
XX Synthetic.
XX Clostridium botulinum.
XX WO9807864-A1.
XX 26-FEB-1998.
XX 22-AUG-1997; 97WO-GB002273.
XX 23-AUG-1996; 96GB-00017671.
XX 13-DEC-1996; 96GB-00025996.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX (SPEY-) SPEYWOOD LAB LTD.
XX Shone CC, Quinn CP, Foster KA;
XX WPI; 1998-169168/15.
XX N-PSDB; AAV26291.
XX Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
XX immunogens or as non-toxic standards for the detection of neurotoxins.
XX Example 1; Page 108-111; 137pp; English.
XX The present sequence represents a recombinant neurotoxin protein from the
XX present invention. The present invention describes recombinant neurotoxin
XX proteins which comprise a first and second domain, where the first domain
XX is adapted to cleave one or more vesicle or plasma-membrane associated
XX proteins essential to exocytosis, and where the second domain is adapted:
XX (a) to translocate the protein into a cell; (b) to increase the
XX solubility of the protein compared to the solubility of the first domain
XX on its own, or (c) both to translocate the protein into a cell and to

CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 871 AA;
Query Match 66.1%; Score 4507; DB 2; Length 871;
Best Local Similarity 99.8%; Pred. No. 1.2e-256;
Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MPFVKQFNKYKDPVNGVDIAIYKIPNAGOMQPVKAFIHNKIWIIPERDFTFTNPEEGDLN 60
Db 1 MQFVKQFNKYKDPVNGVDIAIYKIPNAGOMQPVKAFIHNKIWIIPERDFTFTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYVDSTVLTSTDNKONYLKGVTKLFERIYSTDLGRMLLTISVIRGIPFWGG 120
Db 61 PPPEAKQVPVSYVDSTVLTSTDNKONYLKGVTKLFERIYSTDLGRMLLTISVIRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOFECKSFGEHLVNLTRNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOFECKSFGEHLVNLTRNGY 180
Qy 181 GSTQYIRFSPDFTGPFESLEVDNPLLGAGKATDPAVTLAHELIVAGHRLYGIATNP 240
Db 181 GSTQYIRFSPDFTGPFESLEVDNPLLGAGKATDPAVTLAHELIVAGHRLYGIATNP 240
Qy 241 RVFKVNTNAYYENSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKFDAIATLNKA 300
Db 241 RVFKVNTNAYYENSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKFDAIATLNKA 300
Qy 301 KSVGTGTASLOYKMNVPKEKYLISEDTSGKFSVDKLFKLYKMLTBIYTDNEDNPFVKFKV 360
Db 301 KSVGTGTASLOYKMNVPKEKYLISEDTSGKFSVDKLFKLYKMLTBIYTDNEDNPFVKFKV 360
Qy 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGNLNTNLAANFNQNTENNKNFTKLKNFT 420
Db 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGNLNTNLAANFNQNTENNKNFTKLKNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
Db 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
Qy 481 ITSDTNIEAAENISLDLIQQYVLTFTFDPNEPENISLSDIIGOLELMPNTERFPNG 540
Db 481 ITSDTNIEAAENISLDLIQQYVLTFTFDPNEPENISLSDIIGOLELMPNTERFPNG 540
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 600
Db 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 600
Qy 601 AMELGWVQVYVDFDTSEVSTTDKIADITIIPIYGPALNIGMMLYKDDFVGCALIFSG 660
Db 601 AMELGWVQVYVDFDTSEVSTTDKIADITIIPIYGPALNIGMMLYKDDFVGCALIFSG 660
Qy 661 AVILLEPIEPIAIPVLGTGFALVSIAKVKLTQTIDNALSKEKWKDEVKYIIVTNWLAK 720
Db 661 AVILLEPIEPIAIPVLGTGFALVSIAKVKLTQTIDNALSKEKWKDEVKYIIVTNWLAK 720
Qy 721 VNTQIDILIRKKMEALENQAATKAIINYQVQYTEEEKNNINFNIDLSKKNESINKA 780
Db 721 VNTQIDILIRKKMEALENQAATKAIINYQVQYTEEEKNNINFNIDLSKKNESINKA 780
Qy 781 MININKFLNOCVSYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGQVDRDKDK 840
Db 781 MININKFLNOCVSYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGQVDRDKDK 840
Qy 841 VVNTLSTDIPFQLSKYVDNORLLSTFTTEYIK 871

```

Db      841 VNNLTSDIPFQSKYVDNQRLSTFTTEYIK 871
|||||
RESULT 10
AAW56007
ID AAW56007 standard; protein; 871 AA.
XX
AC AAW56007;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type A LH423/A.
XX
XX Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN W09807864-A1.
XX
XX 26-FEB-1998.
XX
XX 22-AUG-1997; 97WO-GB002273.
XX
XX 23-AUG-1996; 96GB-00017671.
XX
PR 13-DEC-1996; 96GB-00025996.
XX
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX (SPEY-) SPEYWOOD LAB LTD.
XX
XX Shone CC, Quinn CP, Foster KA;
XX
XX WPI; 1998-169168/15.
XX
DR N-PSDB; AAV26279.
XX
XX Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
XX immunogens or as non-toxic standards for the detection of neurotoxins.
XX
XX Example 1; Page 33-35; 137pp; English.
XX
XX The present sequence represents a recombinant neurotoxin protein from the
XX present invention. The present invention describes recombinant neurotoxin
XX proteins which comprise a first and second domain, where the first domain
XX is adapted to cleave one or more vesicle or plasma-membrane associated
XX proteins essential to exocytosis, and where the second domain is adapted:
XX (a) to translocate the protein into a cell; (b) to increase the
XX solubility of the protein compared to the solubility of the first domain
XX on its own, or (c) both to translocate the protein into a cell and to
XX increase the solubility of the protein compared to the solubility of the
XX first domain on its own, the protein being free of clostridial neurotoxin
XX (CN) and free of CN precursor that can be converted into toxin by
XX proteolytic action. The recombinant proteins can be used as therapeutic
XX agents for targeting cells expressing a relevant substrate. The products
XX can also be used as immunogens and as non-toxic standards for the
XX assessment and development of in vitro assays for the detection of
XX functional botulinum or tetanus neurotoxins either in foodstuffs or in
XX environmental samples
XX
XX Sequence 871 AA;
XX
XX Query Match 66.1%; Score 4507; DB 2; Length 871;
XX Best Local Similarity 99.8%; Pred. No. 1.2e-256;
XX Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MPFVNKQFYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 60
XX
XX 1 MQFVNKQFYKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 60
XX
XX 61 PPPEAKQVPVSYDSTYLTSDNEKDNLYLKGVTYKLFERYISTDGLGRMLTSTVIRGIPFWGG 120
XX
XX 61 PPPEAKQVPVSYDSTYLTSDNEKDNLYLKGVTYKLFERYISTDGLGRMLTSTVIRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQPDGYSRSEELNVLIIIGPSADIIIOFECKSGFHEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQPDGYSRSEELNVLIIIGPSADIIIOFECKSGFHEVLNLTNGY 180
QY 181 GSTQYIRFSPDFTFGFEESLEVDTPNLLGAGKFAETDPVTLAHELIIYAGHRLYGTAINEP 240
DB 181 GSTQYIRFSPDFTFGFEESLEVDTPNLLGAGKFAETDPVTLAHELIIYAGHRLYGTAINEP 240
QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDKAFIDSLQENEFRLYYNKFKDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDKAFIDSLQENEFRLYYNKFKDIASTLNKA 300
QY 301 KSIIVGTASLQYMKNVFKKYLSEDTSGKFSVDKLKFDKLYKMLTEIYEDNFVFPFKV 360
DB 301 KSIIVGTASLQYMKNVFKKYLSEDTSGKFSVDKLKFDKLYKMLTEIYEDNFVFPFKV 360
QY 361 LMRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGONTIENNMFTKLKNT 420
DB 361 LMRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGONTIENNMFTKLKNT 420
QY 421 GLFEFYKLLCVRGIIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
DB 421 GLFEFYKLLCVRGIIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
QY 481 ITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIEIENLSSDIIIOCLELMPNIEFPNG 540
DB 481 ITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIEIENLSSDIIIOCLELMPNIEFPNG 540
QY 541 KKYELDKYTMFHYLRAQEFHKGSRITATNSVNEALLNPSRVVTFPSSDYVKKVKATEA 600
DB 541 KKYELDKYTMFHYLRAQEFHKGSRITATNSVNEALLNPSRVVTFPSSDYVKKVKATEA 600
QY 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDFVGALIFSG 660
DB 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDFVGALIFSG 660
QY 661 AVILLEFPIEIPALVIGTALVSIYANKVLTVOTIDNALSKRNEKDEVKYIVTNWLAK 720
DB 661 AVILLEFPIEIPALVIGTALVSIYANKVLTVOTIDNALSKRNEKDEVKYIVTNWLAK 720
QY 721 VNTQIDILRKMKAELENAEATKAIINQYNOYTEEEKNNINFNIDDLSSKLNESINKA 780
DB 721 VNTQIDILRKMKAELENAEATKAIINQYNOYTEEEKNNINFNIDDLSSKLNESINKA 780
QY 781 MININKFLNQCYSVYLMNSMPIYGVKRLDFDASLKDALLKIYIDNRGTILIGQVDRDK 840
DB 781 MININKFLNQCYSVYLMNSMPIYGVKRLDFDASLKDALLKIYIDNRGTILIGQVDRDK 840
QY 841 VNNLTSDIPFQSKYVDNQRLSTFTTEYIK 871
DB 841 VNNLTSDIPFQSKYVDNQRLSTFTTEYIK 871
XX
XX RESULT 11
XX ADL90435
XX ID ADL90435 standard; protein; 871 AA.
XX
XX AC ADL90435;
XX
XX DT 17-JUN-2004 (first entry)
XX
XX DE Clostridial neurotoxin amino acid sequence SEQ ID NO:153.
XX
XX single chain polypeptide; clostridial neurotoxin light chain;
XX clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
XX antibacterial; vaccine; toxin assay; clostridial toxin; detection;
XX botulinum; tetanus.
XX
XX Clostridium botulinum.
XX
XX W02004024909-A2.

```


25-MAR-2004.
 12-SEP-2003; 2003WO-GB003824.
 12-SEP-2002; 2002US-00241596.
 (HEAL-) HEALTH PROTECTION AGENCY.
 Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
 Wayne J;
 WPI; 2004-270039/25.
 N-PSDB; ADL90434.
 New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for developing vaccines against clostridial toxin.
 Claim 1; SEQ ID NO 153; 588pp; English.
 The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.
 Sequence 871 AA;
 Query Match 66.1%; Score 4507; DB 8; Length 871;
 Best Local Similarity 99.8%; Pred. No. 1.2e-256;
 Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 1 MPFVNQFNKDPVNGVDIAIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPPEGDLN 60
 1 MEFVNQFNKDPVNGVDIAIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPPEGDLN 60
 61 PPEAKQVPVSYDYSTYLDNEKONYLKGVTLKERIYSTDIGRMLLTISVIRGIPFWGG 120
 61 PPEAKQVPVSYDYSTYLDNEKONYLKGVTLKERIYSTDIGRMLLTISVIRGIPFWGG 120
 121 STIDTELKVIDTNCINIVIPQGSYRSEELNVIIGPSADIIOPECKSFGEHLNLTNGY 180
 121 STIDTELKVIDTNCINIVIPQGSYRSEELNVIIGPSADIIOPECKSFGEHLNLTNGY 180
 181 GSTQYIRFSDPTFGFPEESLEVDTNPLLGAKPATDPAVTLAHLIYAGHRLYGIAPNP 240
 181 GSTQYIRFSDPTFGFPEESLEVDTNPLLGAKPATDPAVTLAHLIYAGHRLYGIAPNP 240
 241 RVFKVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLANKA 300
 241 RVFKVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLANKA 300
 301 KSIIVGTTASIQYMNQVPEKYLSEDTSGKFSVDKLFKDLKYLKMLTIYEDNPFVKFKV 360
 301 KSIIVGTTASIQYMNQVPEKYLSEDTSGKFSVDKLFKDLKYLKMLTIYEDNPFVKFKV 360

361 LNRKTYLNFEDKAVFKINIVPKVNYITYDGFNLNTLAANFNQONTNINNNFTKLKNFT 420
 361 LNRKTYLNFEDKAVFKINIVPKVNYITYDGFNLNTLAANFNQONTNINNNFTKLKNFT 420
 421 GUFPEFYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVNNWDLPSPSEDFNFTDLNKGEE 480
 421 GUFPEFYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVNNWDLPSPSEDFNFTDLNKGEE 480
 481 ITSNTNIEAAEENISLDLIQQYYLTFNFDNEPENISIEIENLSSDIIGOLELMPNIEPPNG 540
 481 ITSNTNIEAAEENISLDLIQQYYLTFNFDNEPENISIEIENLSSDIIGOLELMPNIEPPNG 540
 541 KKYELDKYTMFHYLRAQEFHKGSRITATNSVNEALLNPSRVYTFPSSDYVKKVKNKATEA 600
 541 KKYELDKYTMFHYLRAQEFHKGSRITATNSVNEALLNPSRVYTFPSSDYVKKVKNKATEA 600
 601 AMPLGHVVEQLVVDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGCALIFSG 660
 601 AMPLGHVVEQLVVDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGCALIFSG 660
 661 AVILLEFIEIAPVLGTGTFALVSIYANKVLTQVTDNALSKRNEKWDVYKYIVTNWLAK 720
 661 AVILLEFIEIAPVLGTGTFALVSIYANKVLTQVTDNALSKRNEKWDVYKYIVTNWLAK 720
 721 VNTQIDILIRKKMKEALENQAETKAIINQYQNYTEEEKNNINFNIDDLSSKLNESINKA 780
 721 VNTQIDILIRKKMKEALENQAETKAIINQYQNYTEEEKNNINFNIDDLSSKLNESINKA 780
 781 MININKFLNQCSYVLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGQVDRLLKDK 840
 781 MININKFLNQCSYVLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGQVDRLLKDK 840
 841 VNNLTSTDIPFQLSKYVDNQRLLSTTFTEYIK 871
 841 VNNLTSTDIPFQLSKYVDNQRLLSTTFTEYIK 871

RESULT 12
 ADL90308
 ID ADL90308 standard; protein; 871 AA.
 XX ADL90308;
 AC
 17-JUN-2004 (first entry)
 DE C. botulinum LH423/A (Q2E, N26K, A27Y) protein SEQ ID NO:26.
 XX single chain polypeptide; clostridial neurotoxin light chain;
 KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
 KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
 KW botulinum; tetanus.
 XX Clostridium botulinum.
 OS
 PN WO2004024909-A2.
 XX
 XX 25-MAR-2004.
 PD
 XX 12-SEP-2003; 2003WO-GB003824.
 PF
 XX 12-SEP-2002; 2002US-00241596.
 PR
 XX (HEAL-) HEALTH PROTECTION AGENCY.
 PA
 XX Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
 PI Wayne J;
 XX WPI; 2004-270039/25.
 DR N-PSDB; ADL90307.
 DR
 XX New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for

developing vaccines against clostridial toxin.

Example 1; SEQ ID NO 26; 588pp; English.

The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

Sequence 871 AA;

Query Match 66.1%; Score 4507; DB 8; Length 871;
Best Local Similarity 99.8%; Pred. No. 1.2e-256;
Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPFVKNQFNYKDPVNGVDIAYIKIPNAGQOPVKAFKHNIWIPIPERDFTTPEEGDLN 60
DB 1 MQFVKNQFNYKDPVNGVDIAYIKIPNAGQOPVKAFKHNIWIPIPERDFTTPEEGDLN 60

QY 61 PPPEAKQVPVSYDSTYLSSTNEKNDYLGKVTYKFERIYSTDLGRMLTISVIRGIPFWGG 120
DB 61 PPPEAKQVPVSYDSTYLSSTNEKNDYLGKVTYKFERIYSTDLGRMLTISVIRGIPFWGG 120

QY 121 STIDTELKVIDTNCINVIQPDGVSYSBELNLVIIGPSADIIQFECKSFGEHVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQPDGVSYSBELNLVIIGPSADIIQFECKSFGEHVLNLTNGY 180

QY 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPVTLAHLIYAGHRLYGIANPN 240
DB 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPVTLAHLIYAGHRLYGIANPN 240

QY 241 RVFKVNTNAYEMSGLEVSPEELTFGHDAKFIDSLQENEFRLYYNKKFDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSPEELTFGHDAKFIDSLQENEFRLYYNKKFDIASTLNKA 300

QY 301 KSIIVGTTASLQYMNQVFKYLLSEDTSKGPSVDKLFKDLKYLKMLTEIYTEDNPFVKFKV 360
DB 301 KSIIVGTTASLQYMNQVFKYLLSEDTSKGPSVDKLFKDLKYLKMLTEIYTEDNPFVKFKV 360

QY 361 LNRKTYLNFDFKAVKINIVPKVNTIYDGNLRTNLAAANFGONTINNNNFTKLKNT 420
DB 361 LNRKTYLNFDFKAVKINIVPKVNTIYDGNLRTNLAAANFGONTINNNNFTKLKNT 420

QY 421 GLPEFYKLLCVRGITTSKTSKLDGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
DB 421 GLPEFYKLLCVRGITTSKTSKLDGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480

QY 481 ITSDTNEAEBNISLDLIQOYYLTFNFDNPEPENISLENSSDIIGOLELMPNIERPENG 540
DB 481 ITSDTNEAEBNISLDLIQOYYLTFNFDNPEPENISLENSSDIIGOLELMPNIERPENG 540

QY 541 KKYELDKYTFHLYRAQEFHGHKGRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
DB 541 KKYELDKYTFHLYRAQEFHGHKGRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600

QY 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
DB 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660

QY 661 AVILLEPIPIAIPVLGTFALVSIANKVLTQVOTIDNALSKRNEKWDDEVKYIIVTNWLAK 720
DB 661 AVILLEPIPIAIPVLGTFALVSIANKVLTQVOTIDNALSKRNEKWDDEVKYIIVTNWLAK 720

QY 721 VNTQIDILIRKMKAELENOAEATKAIINYQVNOYVTEEEKNINFNIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKMKAELENOAEATKAIINYQVNOYVTEEEKNINFNIDDLSSKLNESINKA 780

QY 781 MININKFLNOCSSYLVNMSMPIYGVKRELEDFDASLKDALKYIYDNRGTLLIGQVDRDKDK 840
DB 781 MININKFLNOCSSYLVNMSMPIYGVKRELEDFDASLKDALKYIYDNRGTLLIGQVDRDKDK 840

QY 841 VNNLTSTDIPFQLSKYVDNORLLSTFTTEYIK 871
DB 841 VNNLTSTDIPFQLSKYVDNORLLSTFTTEYIK 871

RESULT 13

ID ADL90284 standard; protein; 871 AA.

XX AC ADL90284;

XX DT 17-JUN-2004 (first entry)

XX DE Clostridium botulinum LH423/A protein SEQ ID NO:2.

XX KW single chain polypeptide; clostridial neurotoxin light chain;
XX KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
XX KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
XX KW botulinum; tetanus.

XX OS Clostridium botulinum.

XX PN WO2004024909-A2.

XX PD 25-MAR-2004.

XX PF 12-SEP-2003; 2003WO-GB003824.

XX PR 12-SEP-2002; 2002US-00241596.

XX PA (HEAL-) HEALTH PROTECTION AGENCY.

XX PI Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
XX PI Wayne J;

XX XX WPI; 2004-270039/25.

XX DR N-PSDB; ADL90283.

XX XX New single chain polypeptides comprising clostridial neurotoxin light and
XX PT heavy chains, useful as positive controls for toxin assays, or for
XX PT developing vaccines against clostridial toxin.

XX PS Example 1; SEQ ID NO 2; 588pp; English.

XX CC The present invention describes a single chain polypeptide comprising
XX CC clostridial neurotoxin light and heavy chains. The single chain
XX CC polypeptide comprises 2 domains: the first domain is a clostridial
XX CC neurotoxin light chain, or its fragment or variant, which is capable of
XX CC cleaving one or more vesicle or plasma membrane associated proteins
XX CC essential to exocytosis; the second domain is a clostridial neurotoxin
XX CC heavy chain H-N portion, or its fragment or variant, which is capable of
XX CC translocating the polypeptide into a cell and/or increasing the
XX CC solubility of the polypeptide compared to the solubility of the first
XX CC domain on its own. The second domain lacks a functional C-terminal part
XX CC of a clostridial neurotoxin heavy chain, designated H-C, which renders
XX CC the polypeptide incapable of binding to cell surface receptors that are

CC the natural cell surface receptors to which native clostridial neurotoxin
CC binds. Also described is a nucleic acid molecule encoding the single
CC chain polypeptide described above. The single chain polypeptide has
CC antibacterial activity, and can be used in vaccines. The single chain
CC polypeptides can be used as positive controls for toxin assays, as
CC reagent components for the synthesis of therapeutic molecules, or for
CC developing vaccines against clostridial toxin. The polypeptides are also
CC useful as non-toxic standards for the assessment and development of in
CC vitro assays for detection of functional botulinum or tetanus neurotoxins
CC in foodstuffs or environmental samples. The present sequence is used in
CC the exemplification of the present invention.
XX
SQ Sequence 871 AA;

Query Match 66.1%; Score 4507; DB 8; Length 871;
Best Local Similarity 99.8%; Pred. No. 1.2e-256;
Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVNKQPNYKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 60
Db 1 MQFVNKQPNYKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHNNKIWIPIERDTFTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYDYSTYLSSTNEDKXNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPPEAKQVPVSYDYSTYLSSTNEDKXNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQPECKSPFGHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQPECKSPFGHEVLNLTNGY 180
Qy 181 GSTQYIRPSPDFTGPFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPN 240
Db 181 GSTQYIRPSPDFTGPFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPN 240
Qy 241 RVPKWNAYYEMSGLEVSPEELTFGCHDAKFTDSIQENEFRLYYNKKFDIASTLNKA 300
Db 241 RVPKWNAYYEMSGLEVSPEELTFGCHDAKFTDSIQENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIIVGTTASLQYMNKVPKEKYLLEDSTSGKFSVDKLFKLYKMLTEIYTEDNFVKFKV 360
Db 301 KSIIVGTTASLQYMNKVPKEKYLLEDSTSGKFSVDKLFKLYKMLTEIYTEDNFVKFKV 360
Qy 361 LNRKTYLNFDFKAVKINIVPKWNTIYDGFNLRNTNLAANFNGQNTINNMFNFKLKQFT 420
Db 361 LNRKTYLNFDFKAVKINIVPKWNTIYDGFNLRNTNLAANFNGQNTINNMFNFKLKQFT 420
Qy 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLRPSSEDFNTDNLNGEE 480
Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLRPSSEDFNTDNLNGEE 480
Qy 481 ITSDTNEAAEENISLDLIQOYLYLTFNFDNEPENISIEENLSSDIIGOLELMPNIERFPNG 540
Db 481 ITSDTNEAAEENISLDLIQOYLYLTFNFDNEPENISIEENLSSDIIGOLELMPNIERFPNG 540
Qy 541 KKYELDKYTFHLYRAQEFHGKGRIALTNSVNEALLNPRVYTFSSDYVKKVKNKATEA 600
Db 541 KKYELDKYTFHLYRAQEFHGKGRIALTNSVNEALLNPRVYTFSSDYVKKVKNKATEA 600
Qy 601 AMFLGWVEQLVYDFTDSTSEVSTTDKADIITIIPIYIGPALNIGMLYKDDFVGCALIFSG 660
Db 601 AMFLGWVEQLVYDFTDSTSEVSTTDKADIITIIPIYIGPALNIGMLYKDDFVGCALIFSG 660
Qy 661 AVILLEPIEIPAIVLGTFFALVSYIANKVLTVQTDNALSKRNEKWDVYKVIYVNNLAK 720
Db 661 AVILLEPIEIPAIVLGTFFALVSYIANKVLTVQTDNALSKRNEKWDVYKVIYVNNLAK 720
Qy 721 VNTQIDILRKQKALENQAETKAIINYQNYQTEEBEKNININIDDLSSKLNESINKA 780
Db 721 VNTQIDILRKQKALENQAETKAIINYQNYQTEEBEKNININIDDLSSKLNESINKA 780
Qy 781 MININKELNOCVSYLMNSMIPYGVKRLDPDASIKDALLKYIYDNRGTLIGQVDRUKDK 840
Db 781 MININKELNOCVSYLMNSMIPYGVKRLDPDASIKDALLKYIYDNRGTLIGQVDRUKDK 840

Qy 841 VNNVLTSDIFPQLSKYVDNORLLSTFTTEYIK 871
Db 841 VNNVLTSDIFPQLSKYVDNORLLSTFTTEYIK 871

RESULT 14

ADL90433
ID ADL90433 standard; protein; 873 AA.

XX AC ADL90433;

DT 17-JUN-2004 (first entry)

XX DE Clostridial neurotoxin amino acid sequence SEQ ID NO:151.

XX KW single chain polypeptide; clostridial neurotoxin light chain;

XX KW Clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;

XX KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;

XX OS Clostridium botulinum.

XX PN W02004024909-A2.

XX PD 25-MAR-2004.

XX PF 12-SEP-2003; 2003WO-GB003824.

XX PR 12-SEP-2002; 2002US-00241596.

XX PA (HEAL-) HEALTH PROTECTION AGENCY.

XX PI Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;

XX PI Wayne J;

XX DR WPI; 2004-270039/25.

XX DR N-PSDB; ADL90432.

XX New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for developing vaccines against clostridial toxin.

PS Claim 1; SEQ ID NO 151; 588pp; English.

XX The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

SQ Sequence 873 AA;

Query Match 66.1%; Score 4507; DB 8; Length 873;
Best Local Similarity 99.8%; Pred. No. 1.2e-256;

Matches		869;	Conservative	1;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	1	MPFVNKQPNYKDPVNGVDIAIYIKIPNAGQOPVKA	FIHKNKIHWIPIERDFTFTPEGDNL	60						
Db	3	MEFVNKQPNYKDPVNGVDIAIYIKIPNAGQOPVKA	FIHKNKIHWIPIERDFTFTPEGDNL	62						
Qy	61	PPPEAKQVPVSYDYDSTYLSSTNEKNDYLGKVT	KLFIERYSTDLGRMLLTSIVRGIPFWGG	120						
Db	63	PPPEAKQVPVSYDYDSTYLSSTNEKNDYLGKVT	KLFIERYSTDLGRMLLTSIVRGIPFWGG	122						
Qy	121	STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPS	ADIIQFECKSFGEHVLNLTNGY	180						
Db	123	STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPS	ADIIQFECKSFGEHVLNLTNGY	182						
Qy	181	GSTQYIRFSDPFTFGFPESELEVDNPLLGAGKAT	DPAVTLAHELIYAGHRLYGIAINPN	240						
Db	183	GSTQYIRFSDPFTFGFPESELEVDNPLLGAGKAT	DPAVTLAHELIYAGHRLYGIAINPN	242						
Qy	241	RVFKVNTNAYYENSGLEVSFEELRTFGHDAKFT	DSIQENEFRLYYNKFDAIASTLKA	300						
Db	243	RVFKVNTNAYYENSGLEVSFEELRTFGHDAKFT	DSIQENEFRLYYNKFDAIASTLKA	302						
Qy	301	KSIIVGTTASIQYKQNVKPKYLLSEDTSKGFSV	DKLKFYKMLTBIYTEDNFVFPKV	360						
Db	303	KSIIVGTTASIQYKQNVKPKYLLSEDTSKGFSV	DKLKFYKMLTBIYTEDNFVFPKV	362						
Qy	361	LNKTYLNEFKAVFKINIVPKVNTIYDGFNLRNT	LNLAANFNGONTNINNNFTKLKQFT	420						
Db	363	LNKTYLNEFKAVFKINIVPKVNTIYDGFNLRNT	LNLAANFNGONTNINNNFTKLKQFT	422						
Qy	421	GLFFFYKLLCVRGIIITSKTSKLDGYNKALNDL	CIKVNNWDLFPSPSEDNFTDNLKGE	480						
Db	423	GLFFFYKLLCVRGIIITSKTSKLDGYNKALNDL	CIKVNNWDLFPSPSEDNFTDNLKGE	482						
Qy	481	ITSDTNEAABENISLDLIQOYYLTFNFDNEPEN	ISNIENLSSDIIGOLELMPNTERPFG	540						
Db	483	ITSDTNEAABENISLDLIQOYYLTFNFDNEPEN	ISNIENLSSDIIGOLELMPNTERPFG	542						
Qy	541	KYELDKYTMFHYLRAQEFHGKRIALNTSNVNEA	LLNPSRVYTFSSDYVKKVKNKATEA	600						
Db	543	KYELDKYTMFHYLRAQEFHGKRIALNTSNVNEA	LLNPSRVYTFSSDYVKKVKNKATEA	602						
Qy	601	AMFLGWEQLVYDFTDETSEVSTTDKIADITII	PIYIGPALNIGNMLYKDDFVGALLFSG	660						
Db	603	AMFLGWEQLVYDFTDETSEVSTTDKIADITII	PIYIGPALNIGNMLYKDDFVGALLFSG	662						
Qy	661	AVILLEFIPEAIPIVLGTFALVSIYANKVLTVQ	TIDNALSKRNEKWEVYKIIVTNWLAK	720						
Db	663	AVILLEFIPEAIPIVLGTFALVSIYANKVLTVQ	TIDNALSKRNEKWEVYKIIVTNWLAK	722						
Qy	721	VNTQIDLRKWKKEALENQAEATKAIINYQYNQY	TEBEKNNINFNIDLSKLNESINKA	780						
Db	723	VNTQIDLRKWKKEALENQAEATKAIINYQYNQY	TEBEKNNINFNIDLSKLNESINKA	782						
Qy	781	MININKFLNQCSVSLNMSMIPYGVKRLIEDFAS	LKDALLKYIYDNRGTLIGQVDRLKDK	840						
Db	783	MININKFLNQCSVSLNMSMIPYGVKRLIEDFAS	LKDALLKYIYDNRGTLIGQVDRLKDK	842						
Qy	841	VNNTLSTDIPQLSKYVDNQRLSTFTTEYIK	871							
Db	843	VNNTLSTDIPQLSKYVDNQRLSTFTTEYIK	873							
RESULT 15										
ID	ADL90431	standard; protein; 873 AA.								
XX	AC	ADL90431;								
XX	DT	17-JUN-2004 (first entry)								
XX	DE	Clostridium neurotoxin amino acid sequence SEQ ID NO:149.								

KW	single chain polypeptide; clostridial neurotoxin light chain;
KW	clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW	antibacterial; vaccine; toxin assay; clostridial toxin; detection;
XX	botulinum; tetanus.
OS	Clostridium botulinum.
XX	WO2004024909-A2.
PN	25-MAR-2004.
XX	12-SEP-2003; 2003WO-GB003824.
PF	12-SEP-2002; 2002US-00241596.
XX	(HEAL-) HEALTH PROTECTION AGENCY.
PA	Shone CC, Poster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
PI	Wayne J;
PI	WPI; 2004-270039/25.
XX	N-PSDB; ADL90430.
DR	New single chain polypeptides comprising clostridial neurotoxin light and
XX	heavy chains, useful as positive controls for toxin assays, or for
PT	developing vaccines against clostridial toxin.
PT	Claim 1; SEQ ID NO 149; 588pp; English.
PS	The present invention describes a single chain polypeptide comprising
XX	clostridial neurotoxin light and heavy chains. The single chain
CC	polypeptide comprises 2 domains: the first domain is a clostridial
CC	neurotoxin light chain, or its fragment or variant, which is capable of
CC	cleaving one or more vesicle or plasma membrane associated proteins
CC	essential to exocytosis; the second domain is a clostridial neurotoxin
CC	heavy chain H-N portion, or its fragment or variant, which is capable of
CC	translocating the polypeptide into a cell and/or increasing the
CC	solubility of the polypeptide compared to the solubility of the first
CC	domain on its own. The second domain lacks a functional C-terminal part
CC	of a clostridial neurotoxin heavy chain, designated H-C, which renders
CC	the polypeptide incapable of binding to cell surface receptors that are
CC	the natural cell surface receptors to which native clostridial neurotoxin
CC	binds. Also described is a nucleic acid molecule encoding the single
CC	chain polypeptide described above. The single chain polypeptide has
CC	antibacterial activity, and can be used in vaccines. The single chain
CC	polypeptides can be used as positive controls for toxin assays, as
CC	reagent components for the synthesis of therapeutic molecules, or for
CC	developing vaccines against clostridial toxin. The polypeptides are also
CC	useful as non-toxic standards for the assessment and development of in
CC	vitro assays for detection of functional botulinum or tetanus neurotoxins
CC	in foodstuffs or environmental samples. The present sequence is used in
CC	the exemplification of the present invention.
XX	Sequence 873 AA;
SQ	Query Match 66.1%; Score 4507; DB 8; Length 873;
	Best Local Similarity 99.8%; Pred. No. 1.2e-256;
	Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy	1 MPFVNKQPNYKDPVNGVDIAIYIKIPNAGQOPVKA
Db	3 MEFVNKQPNYKDPVNGVDIAIYIKIPNAGQOPVKA
Qy	61 PPPEAKQVPVSYDYDSTYLSSTNEKNDYLGKVT
Db	63 PPPEAKQVPVSYDYDSTYLSSTNEKNDYLGKVT
Qy	121 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPS
Db	123 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPS
Qy	181 GSTQYIRFSDPFTFGFPESELEVDNPLLGAGKAT
XX	DPAVTLAHELIYAGHRLYGIAINPN 240

Db	183	GSTQYIRFSDFTFGFESLEVDNPLLGAGKFAATDPAVTLAHELIIHAGHRLYGIAINPN	242
Qy	241	RVFKVNTNAYEMSGLEVSPEELRTFGCHDAKFIDSIQENEFRLYYNKKFKDIASTLNKA	300
Db	243	RVFKVNTNAYEMSGLEVSPEELRTFGCHDAKFIDSIQENEFRLYYNKKFKDIASTLNKA	302
Qy	301	KSIVGTTASLOYMKQNVFKKYLISEDTSGKFSVDKLFKLYKMLTEIYTEDNPFVKFKV	360
Db	303	KSIVGTTASLOYMKQNVFKKYLISEDTSGKFSVDKLFKLYKMLTEIYTEDNPFVKFKV	362
Qy	361	LNKTYLNFDPKAVFKINIYVKNVYTIYDGFNLRNTNLAANFNQONTENNNTFKLNFT	420
Db	363	LNKTYLNFDPKAVFKINIYVKNVYTIYDGFNLRNTNLAANFNQONTENNNTFKLNFT	422
Qy	421	GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE	480
Db	423	GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE	482
Qy	481	ITSDTNIEAAEENISLDLIQOYILTFFNFDNEPENISIENTLSSDIIGQLELMPNIERFNG	540
Db	483	ITSDTNIEAAEENISLDLIQOYILTFFNFDNEPENISIENTLSSDIIGQLELMPNIERFNG	542
Qy	541	KYELDKYTWPHYLRAQEFHKGSRIALTNSVNEALNPSRVYTFPSSDVYKVNKATEA	600
Db	543	KYELDKYTWPHYLRAQEFHKGSRIALTNSVNEALNPSRVYTFPSSDVYKVNKATEA	602
Qy	601	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG	660
Db	603	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG	662
Qy	661	AVILLEPIPEIAPVLGTFALVSIANKVLTVOQIDNALSKRNEKWDDEVYKIVTNWLAK	720
Db	663	AVILLEPIPEIAPVLGTFALVSIANKVLTVOQIDNALSKRNEKWDDEVYKIVTNWLAK	722
Qy	721	VNTQIDLIIRKKMEALENQAEATKAIINYQYNOYTEEEKNNINFNIDDLSSKLNESINKA	780
Db	723	VNTQIDLIIRKKMEALENQAEATKAIINYQYNOYTEEEKNNINFNIDDLSSKLNESINKA	782
Qy	781	MININKFLNQCYSVYLMNSMIPYGVKREDFDASLKDALLKYIYDNRGTLLIGQVDRLKDK	840
Db	783	MININKFLNQCYSVYLMNSMIPYGVKREDFDASLKDALLKYIYDNRGTLLIGQVDRLKDK	842
Qy	841	VNNTLSTDIPFQLSKYVDNQRLLSFTTEYIK	871
Db	843	VNNTLSTDIPFQLSKYVDNQRLLSFTTEYIK	873

Search completed: January 31, 2005, 14:04:11
Job time : 182 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:03 ; Search time 43 Seconds
(without alignments)
2008.047 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFWKQFNKDPVNGVDIA.....EFIPVDGCGERPLHHHH 1302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6756	99.0	1296	1	US-08-480-604A-28
2	6756	99.0	1296	2	US-08-405-496A-28
3	6756	99.0	1296	3	US-08-915-136-28
4	6756	99.0	1296	4	US-09-084-517-28
5	4515	66.2	871	4	US-09-255-829-8
6	4507	66.1	871	4	US-09-255-829-2
7	4507	66.1	871	4	US-09-255-829-26
8	4501	66.0	1013	4	US-09-255-829-18
9	4495	65.9	873	4	US-09-255-829-6
10	4495	65.9	875	4	US-09-255-829-10
11	4495	65.9	894	4	US-09-255-829-4
12	4494.5	65.9	907	4	US-09-255-829-16
13	4494.5	65.9	953	4	US-09-255-829-14
14	4493.5	65.9	878	4	US-09-255-829-12
15	2348	34.4	448	4	US-09-288-326A-7
16	2307	33.8	438	1	US-08-480-604A-23
17	2307	33.8	438	2	US-08-405-496A-23
18	2307	33.8	438	3	US-08-915-136-23
19	2307	33.8	438	4	US-09-084-517-23
20	2307	33.8	462	1	US-08-480-604A-26
21	2307	33.8	462	2	US-08-405-496A-26
22	2307	33.8	462	3	US-08-915-136-26
23	2307	33.8	462	4	US-09-084-517-26
24	2167	31.8	423	4	US-09-288-326A-8
25	2146.5	31.5	1169	4	US-09-255-829-20
26	2036	29.8	382	4	US-09-288-326A-9
27	1892.5	27.7	1315	3	US-08-913-880C-1

28	1451.5	21.3	858	4	US-09-255-829-22	Sequence 22, Appl
29	1451.5	21.3	858	4	US-09-255-829-29	Sequence 23, Appl
30	1348	19.8	874	3	US-08-913-880C-11	Sequence 11, Appl
31	1348	19.8	875	3	US-08-913-880C-10	Sequence 10, Appl
32	1345	19.7	862	3	US-08-913-880C-14	Sequence 14, Appl
33	1345	19.7	865	3	US-08-913-880C-13	Sequence 13, Appl
34	1345	19.7	866	3	US-08-913-880C-12	Sequence 12, Appl
35	1343	19.7	853	3	US-08-913-880C-17	Sequence 17, Appl
36	1343	19.7	858	3	US-08-913-880C-16	Sequence 16, Appl
37	1343	19.7	860	3	US-08-913-880C-15	Sequence 15, Appl
38	1071	15.7	207	3	US-08-446-114A-21	Sequence 21, Appl
39	733	10.7	141	4	US-09-465-276-1	Sequence 1, Appl
40	727	10.7	140	3	US-08-446-114A-22	Sequence 22, Appl
41	653.5	9.6	618	1	US-08-668-381A-5	Sequence 5, Appl
42	646.5	9.5	548	4	US-09-255-829-24	Sequence 24, Appl
43	634.5	9.3	452	1	US-07-618-312A-2	Sequence 2, Appl
44	634.5	9.3	452	1	US-08-280-228-2	Sequence 2, Appl
45	633.5	9.3	452	1	US-07-618-312A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 1296 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-480-604A-28

Query Match 99.0%; Score 6756; DB 1; Length 1296;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVNKQENYKDPVNGVDIAIYIKIPNAGOMQPVAKFHKHNIWIIPERDFTNPEEGDLN 60
DB 1 MQFVNKQENYKDPVNGVDIAIYIKIPNAGOMQPVAKFHKHNIWIIPERDFTNPEEGDLN 60
QY 61 PPEAKQVPVSYDYSTYLDSTNEKDNLYKGYTKLPERIYSTDLGRMLLTSIVRGI PFWGG 120
DB 61 PPEAKQVPVSYDYSTYLDSTNEKDNLYKGYTKLPERIYSTDLGRMLLTSIVRGI PFWGG 120
QY 121 STIDTELKVIDTNCINVIQDPSYRSEELNLVIIGPSADIIIOFECKSFGEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQDPSYRSEELNLVIIGPSADIIIOFECKSFGEVLNLTNGY 180
QY 181 GSTQYIRPSPDFTGFEESELDVNTNPLLGAGKFPATDPAVTLAHELIYAGHRLYGIAPNP 240
DB 181 GSTQYIRPSPDFTGFEESELDVNTNPLLGAGKFPATDPAVTLAHELIYAGHRLYGIAPNP 240
QY 241 RVFKVNTNAYYEMSGLEVSPEELTFGGHDAKFIDSLOENEFRLYYNKKPKDIASTLNKA 300
DB 241 RVFKVNTNAYYEMSGLEVSPEELTFGGHDAKFIDSLOENEFRLYYNKKPKDIASTLNKA 300
QY 301 KSIIVGTTASLOYKMNFKPKYLLSEDTSGKFSVDKLFKPKLYKMLTEIYTEDNFVKPFKV 360
DB 301 KSIIVGTTASLOYKMNFKPKYLLSEDTSGKFSVDKLFKPKLYKMLTEIYTEDNFVKPFKV 360
QY 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGFNLRTNLAANFGONTENNMFYKLNFT 420
DB 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGFNLRTNLAANFGONTENNMFYKLNFT 420
QY 421 GLFEFYKLLCVRGITTSKTSLDGYNKALNDLCIKVNNWDLFPSSEDFNTDNLKGE 480
DB 421 GLFEFYKLLCVRGITTSKTSLDGYNKALNDLCIKVNNWDLFPSSEDFNTDNLKGE 480
QY 481 ITSDTNEAABENISLDLIQYLYTFNPDNEPENISIENTSSDIIGOLELMPNTERFPNG 540
DB 481 ITSDTNEAABENISLDLIQYLYTFNPDNEPENISIENTSSDIIGOLELMPNTERFPNG 540
QY 541 KKYELDKYTMPHYLRQAQEFHGKSRIALTNSVNEALLNPSRYVTFSSDYVKKYKNAKATEA 600
DB 541 KKYELDKYTMPHYLRQAQEFHGKSRIALTNSVNEALLNPSRYVTFSSDYVKKYKNAKATEA 600
QY 601 AMFLGWEOQLVYDFTDETSEVSTTDKADIITIIPIYGPALNIGMLYKODFVGCALIFSG 660
DB 601 AMFLGWEOQLVYDFTDETSEVSTTDKADIITIIPIYGPALNIGMLYKODFVGCALIFSG 660
QY 661 AVILLEFPEITAPVLGTFALVSIVANKVLTQVTDNALSKRNEKWDVEVYKIVTNWLAK 720
DB 661 AVILLEFPEITAPVLGTFALVSIVANKVLTQVTDNALSKRNEKWDVEVYKIVTNWLAK 720
QY 721 VNTQIDILIRKQKALENOAATKAIINYOQNOVTESEKKNINENIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKQKALENOAATKAIINYOQNOVTESEKKNINENIDDLSSKLNESINKA 780
QY 781 MININKFNLQCSVSYLWNSMIPYGVKRLLEDPAKDLAKLYIYDNRGTGIGQVDRLLKDK 840
DB 781 MININKFNLQCSVSYLWNSMIPYGVKRLLEDPAKDLAKLYIYDNRGTGIGQVDRLLKDK 840
QY 841 VNNTLSTDIPFQLSKYVDNQRLLSFTTEYIKNIINTSILNRYESNHLIDLRSYASKINI 900

DB 841 VNNTLSTDIPFQLSKYVDNQRLLSFTTEYIKNIINTSILNRYESNHLIDLRSYASKINI 900
QY 901 GSKVNPDPIDKNOIQLFNLESSKIEVLKNAIVVNSMYENFSTFWIRIPKYFNSISLNN 960
DB 901 GSKVNPDPIDKNOIQLFNLESSKIEVLKNAIVVNSMYENFSTFWIRIPKYFNSISLNN 960
QY 961 EYTIINCWNNNSGKWSLNYGELIITWLTQDTEIKQKRVVFKYSQMINISDIYNRWIFVTIT 1020
DB 961 EYTIINCWNNNSGKWSLNYGELIITWLTQDTEIKQKRVVFKYSQMINISDIYNRWIFVTIT 1020
QY 1021 NNRLLNSKIYINGRLIDQKPI SNLGNTHASNNIMFKLDGCRDTHRYIWIYFNLPDKELN 1080
DB 1021 NNRLLNSKIYINGRLIDQKPI SNLGNTHASNNIMFKLDGCRDTHRYIWIYFNLPDKELN 1080
QY 1081 EKEIKDYDNQNSGILKDFWGDYLOVDKPYMLNLYDPNKYVDVNNVVGIRGYMYLKGPR 1140
DB 1081 EKEIKDYDNQNSGILKDFWGDYLOVDKPYMLNLYDPNKYVDVNNVVGIRGYMYLKGPR 1140
QY 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
DB 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
QY 1201 GVEKILSALSIIPDVGNLSQVVMKSKNDQGITNCKKRNLODNGNDIGTGFHGFNNIAK 1260
DB 1201 GVEKILSALSIIPDVGNLSQVVMKSKNDQGITNCKKRNLODNGNDIGTGFHGFNNIAK 1260
QY 1261 LVASNNWNRQIERSSRTLGCSEWEPFIPVDDGWERPL 1296
DB 1261 LVASNNWNRQIERSSRTLGCSEWEPFIPVDDGWERPL 1296

RESULT 2

US-08-405-496A-28

Sequence 28, Application US/08405496A

Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.

TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

TITLE OF INVENTION: NEUROTOXIN

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/405,496A

APPLICATION NUMBER: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01308

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-28

Query Match 99.0%; Score 6756; DB 2; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MPFVKQFNKYDVPNGVDIAIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTFTNPEEGDLN 60
DB 1 MQFVKQFNKYDVPNGVDIAIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTFTNPEEGDLN 60
QY 61 PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
DB 61 PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQDGSYSRSEELNLIIGPSADIIQPECKSFGEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQDGSYSRSEELNLIIGPSADIIQPECKSFGEVLNLTNGY 180
QY 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATLAHELIIYAGHRLYGIAINPN 240
DB 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATLAHELIIYAGHRLYGIAINPN 240
QY 241 RVFKVNTNAYEMSGLEVSFEELTFGHDADFTDSLOENEFRLYYNKKFKDIASLTNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELTFGHDADFTDSLOENEFRLYYNKKFKDIASLTNKA 300
QY 301 KSIQVGTASLOYMKVNFKEKYLSEDTSGKFSVDKLFKLYKMLTEIYEDNPFVKFKV 360
DB 301 KSIQVGTASLOYMKVNFKEKYLSEDTSGKFSVDKLFKLYKMLTEIYEDNPFVKFKV 360
QY 361 LNRKTYLNFKAQFKINIVPKVNTIYDGFNLRLNLAANFNGONTINNNTFKLNFT 420
DB 361 LNRKTYLNFKAQFKINIVPKVNTIYDGFNLRLNLAANFNGONTINNNTFKLNFT 420
QY 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEE 480
DB 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEE 480
QY 481 ITSNTNEAENISLDIIQYLYLTFNPDNEPENISNIENSSDIIGOLELMPNIERPNG 540
DB 481 ITSNTNEAENISLDIIQYLYLTFNPDNEPENISNIENSSDIIGOLELMPNIERPNG 540
QY 541 KYELDKYTWPHYLRAOEFHGKSRIALTNSVNEALLNPKSRVYTFSSDYVKKVKNKATEA 600
DB 541 KYELDKYTWPHYLRAOEFHGKSRIALTNSVNEALLNPKSRVYTFSSDYVKKVKNKATEA 600
QY 601 AMFLGWQEQLVYDFTDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDFVGALIFSG 660
DB 601 AMFLGWQEQLVYDFTDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDFVGALIFSG 660
QY 661 AVILLEPIPEIATPVLGTFAVSIANKVLTVQIDNALSKRNEKWDVEVYKIIVTNWLAK 720
DB 661 AVILLEPIPEIATPVLGTFAVSIANKVLTVQIDNALSKRNEKWDVEVYKIIVTNWLAK 720
QY 721 VNTQIDILIRKQWKEALNQAEATKAIINYOQVTEEEKNNINFIIDLSKKNESINKA 780
DB 721 VNTQIDILIRKQWKEALNQAEATKAIINYOQVTEEEKNNINFIIDLSKKNESINKA 780
QY 781 MININKFLNQCVSYLNMNMPYGVRLDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
DB 781 MININKFLNQCVSYLNMNMPYGVRLDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
QY 841 VNTLSTDIPFQLSKYVDNQRLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI 900

DB 841 VNTLSTDIPFQLSKYVDNQRLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI 900
QY 901 GSKVNFDPIDKNQIQLFNLESSKIEVLKNAIVNMYENFSTFSWIRIPKYFNISLNN 960
DB 901 GSKVNFDPIDKNQIQLFNLESSKIEVLKNAIVNMYENFSTFSWIRIPKYFNISLNN 960
QY 961 EYTIINCMMNSGKVSILNYGEIITWLODQEIQRVVFYKYSQMINISDYINRWIFVTIT 1020
DB 961 EYTIINCMMNSGKVSILNYGEIITWLODQEIQRVVFYKYSQMINISDYINRWIFVTIT 1020
QY 1021 NNRLLNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
DB 1021 NNRLLNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
QY 1081 EKEIKOLDYDQNSGILKDFWGDYLYQDKPYMLNLYDPNKKYVDVNNVIRGMYLKGPR 1140
DB 1081 EKEIKOLDYDQNSGILKDFWGDYLYQDKPYMLNLYDPNKKYVDVNNVIRGMYLKGPR 1140
QY 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKONIVRNDRVYINVVKNKESYRLATNASQA 1200
DB 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKONIVRNDRVYINVVKNKESYRLATNASQA 1200
QY 1201 GVEKILSALREIPDVGNLQVVMKSNQDQITNKKCNLQDNNNGNDIGFIFGHFNNAIAK 1260
DB 1201 GVEKILSALREIPDVGNLQVVMKSNQDQITNKKCNLQDNNNGNDIGFIFGHFNNAIAK 1260
QY 1261 LVASNNYNRQIERSRSLTGCSEWFIPIVDDGWMGERPL 1296
DB 1261 LVASNNYNRQIERSRSLTGCSEWFIPIVDDGWMGERPL 1296

RESULT 3

US-08-915-136-28
Sequence 28, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHAY, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELEPHONE: (415) 703-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-28

Query Match 99.0%; Score 6756; DB 3; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVNKQNYKDPVNGVDIAYIKIPNAGQMPKAFKIHNIKIWIPIERDTFTNPEEGDLN 60
DB 1 MQFVNKQNYKDPVNGVDIAYIKIPNAGQMPKAFKIHNIKIWIPIERDTFTNPEEGDLN 60
QY 61 PPPEAKQPVSYDYSTYLTSTNEKDNLYKGYTKLPERIYSTDGLRMLLTSIVRGIPFWGG 120
DB 61 PPPEAKQPVSYDYSTYLTSTNEKDNLYKGYTKLPERIYSTDGLRMLLTSIVRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIOPGVSRSBELNVIIGPSADIIOFCKSGFGHEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIOPGVSRSBELNVIIGPSADIIOFCKSGFGHEVLNLTNGY 180
QY 181 GSTQYIRFSPDFTGFEBSLEVDNPLLGAGKATDPAVTLAHELIIYAGHRLYGIAPNP 240
DB 181 GSTQYIRFSPDFTGFEBSLEVDNPLLGAGKATDPAVTLAHELIIYAGHRLYGIAPNP 240
QY 241 RVFKVNTNAYEMSGLEVSPELFTFGCHDAKFTDSLQENEFRLYYNKKFDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSPELFTFGCHDAKFTDSLQENEFRLYYNKKFDIASTLNKA 300
QY 301 KSIIVGTTASLYQMNKVFKEKLLSGEDTSKGSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
DB 301 KSIIVGTTASLYQMNKVFKEKLLSGEDTSKGSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
QY 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNNTFKLNFT 420
DB 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNNTFKLNFT 420
QY 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKEE 480
DB 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKEE 480
QY 481 ITSDTNTAEENISLDLIQOYLYITFNPDNENISLENSSDIIGOLELMPNTERPPNG 540
DB 481 ITSDTNTAEENISLDLIQOYLYITFNPDNENISLENSSDIIGOLELMPNTERPPNG 540
QY 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNBALLNPSRVYTFSSDYVKKYNKATEA 600
DB 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNBALLNPSRVYTFSSDYVKKYNKATEA 600
QY 601 AMFLGWELQVYDFTDSEVSTTDKIADITIIIPYGPALNIGNMLYKDDFVGLIFSG 660
DB 601 AMFLGWELQVYDFTDSEVSTTDKIADITIIIPYGPALNIGNMLYKDDFVGLIFSG 660
QY 661 AVILLEFPEIAPVLGTFALVSIVANKVLTVQTDNALSRRNEKWDVEVYKIVTNWLAK 720
DB 661 AVILLEFPEIAPVLGTFALVSIVANKVLTVQTDNALSRRNEKWDVEVYKIVTNWLAK 720
QY 721 VNTQIDLIRKKWKEALENQAEATKAIINYQVQTEBEKKNINFNIDDLSSKLNESINKA 780

Db 721 VNTQIDLIRKKWKEALENQAEATKAIINYQVQTEBEKKNINFNIDDLSSKLNESINKA 780
QY 781 MININKELNOCVSYLYMNSMIPYVKELEDASLKDALLKYIYDNEGTILGOVDRUKDK 840
Db 781 MININKELNOCVSYLYMNSMIPYVKELEDASLKDALLKYIYDNEGTILGOVDRUKDK 840
QY 841 VVNTLSTDIPFQLSKYVDNORLLSTFTYEIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Db 841 VVNTLSTDIPFQLSKYVDNORLLSTFTYEIKNIINTSILNLRYESNHLIDLSRYASKINI 900
QY 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVNMYENFSTFWIRIPKYNFNSILNN 960
Db 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVNMYENFSTFWIRIPKYNFNSILNN 960
QY 961 EYTIINCWENNSGKVSILNYGEIITWLODQTEIKQORVVVKYSOMINISDYINRWIFVTIT 1020
Db 961 EYTIINCWENNSGKVSILNYGEIITWLODQTEIKQORVVVKYSOMINISDYINRWIFVTIT 1020
QY 1021 NNRLNNSKIYINGRLIDQKPI SNLGNHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
Db 1021 NNRLNNSKIYINGRLIDQKPI SNLGNHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
QY 1081 EKEIKDLYDNQSGSILKDPWGDYLOYDKPYMMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Db 1081 EKEIKDLYDNQSGSILKDPWGDYLOYDKPYMMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
QY 1141 GSVMTTNIYLNSSILYRGTKFIIKKYASGNKDNIVRNNDRVYINVVKNKEVRLATNASQA 1200
Db 1141 GSVMTTNIYLNSSILYRGTKFIIKKYASGNKDNIVRNNDRVYINVVKNKEVRLATNASQA 1200
QY 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNKKCNLQDNNGNDIGFIFGHQFNNAK 1260
Db 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNKKCNLQDNNGNDIGFIFGHQFNNAK 1260
QY 1261 LVASNTWNTROIESRRTLCGSWEFIPVDDGWERPL 1296
Db 1261 LVASNTWNTROIESRRTLCGSWEFIPVDDGWERPL 1296

RESULT 4
US-09-084-517-28
Sequence 28, Application US/09084517
Patent No. 6613329
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESSES:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/084,517
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907		661	AVILLEPIEAI	PVLGTFALVSYIANKVLTVOTIDNLSKRNEKWDVYKIVTNWLAK	720
FILING DATE: 02-DEC-1993		Qy	721	VNTQIDILIRKMKKEALENQAEATKAIINYQYNOYQYTEEKNNINFNIDDLSSKLNESINKA	780
PRIOR APPLICATION DATA:		Dd	721	VNTQIDILIRKMKKEALENQAEATKAIINYQYNOYQYTEEKNNINFNIDDLSSKLNESINKA	780
FILING DATE: 04-DEC-1992		Qy	781	MININFLANQCSVSYLMNSMIPYGVKRELEDFDASLKDALLKYIDNRGTLIGQVDRUKDK	840
PRIOR APPLICATION DATA:		Dd	781	MININFLANQCSVSYLMNSMIPYGVKRELEDFDASLKDALLKYIDNRGTLIGQVDRUKDK	840
FILING DATE: 31-OCT-1989		Qy	841	VNNTLSTDIPOFQSKVVDNORLLSTFTEYIKNTINTSILNLRYESNHLIDLRSYASIKINI	900
ATTORNEY/AGENT INFORMATION:		Dd	841	VNNTLSTDIPOFQSKVVDNORLLSTFTEYIKNTINTSILNLRYESNHLIDLRSYASIKINI	900
NAME: CARROLL, PETER G.		Qy	901	GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVNMYENFSTFWIRIPKYFNISILNN	960
REGISTRATION NUMBER: 32,837		Dd	901	GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVNMYENFSTFWIRIPKYFNISILNN	960
REFERENCE/DOCKET NUMBER: OPND-01610		Qy	961	EYTIINCMMNSGKVSILNYGEBIIWTLQDTEIKQKRVVFKYSOMINISDYINRWIFVTIT	1020
TELECOMMUNICATION INFORMATION:		Dd	961	EYTIINCMMNSGKVSILNYGEBIIWTLQDTEIKQKRVVFKYSOMINISDYINRWIFVTIT	1020
TELEPHONE: (415) 705-8410		Qy	1021	NNRLNNSKIYINGRLIDOKPISNLGNHASNIMFVKLDGCRDTHRYIWKYFNLFDKELN	1080
TELEFAX: (415) 397-8338		Dd	1021	NNRLNNSKIYINGRLIDOKPISNLGNHASNIMFVKLDGCRDTHRYIWKYFNLFDKELN	1080
INFORMATION FOR SEQ ID NO: 28:		Qy	1081	EKEIKOLDYDQNSGILKDFWGDYQYDQYKPYMLNLYDPNKYVDVNNVGVIRGYMYLKGPR	1140
SEQUENCE CHARACTERISTICS:		Dd	1081	EKEIKOLDYDQNSGILKDFWGDYQYDQYKPYMLNLYDPNKYVDVNNVGVIRGYMYLKGPR	1140
LENGTH: 1296 amino acids		Qy	1141	GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINNVVQNKYRLATNASQA	1200
TYPE: amino acid		Dd	1141	GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINNVVQNKYRLATNASQA	1200
TOPOLOGY: linear		Qy	1201	GVSKILSALRIPDVGNLSQVVMKSKNDQGITNKKCNALDNNNGNDIGFIFGHOFNNIAK	1260
MOLECULE TYPE: protein		Dd	1201	GVSKILSALRIPDVGNLSQVVMKSKNDQGITNKKCNALDNNNGNDIGFIFGHOFNNIAK	1260
US-09-084-517-28		Qy	1261	LVASNWNRIERSRSLGCSWEFIPVDDGWERPL	1296
		Dd	1261	LVASNWNRIERSRSLGCSWEFIPVDDGWERPL	1296
Query Match		99.0%; Score 6756; DB 4; Length 1296;			
Best Local Similarity		99.8%; Pred. No. 0;			
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;					
Qy	1	MPFVKQFNKDPVNGVDIAVILIPNAGQMPVKAFKIHKNKIWIPIERDFTFNPDEGDLN	60		
Dd	1	MQVNNKQFNKDPVNGVDIAVILIPNAGQMPVKAFKIHKNKIWIPIERDFTFNPDEGDLN	60		
Qy	61	PPPEAKQVPVSYDYSTYLSLTDNEKNDYLGKVTLEFRIYSTDGLRMLLTSIVRGIPFWGG	120		
Dd	61	PPPEAKQVPVSYDYSTYLSLTDNEKNDYLGKVTLEFRIYSTDGLRMLLTSIVRGIPFWGG	120		
Qy	121	STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY	180		
Dd	121	STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY	180		
Qy	181	GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIIYAGHRLYGIANPN	240		
Dd	181	GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIIYAGHRLYGIANPN	240		
Qy	241	RVFKVNTNAYEMSGLEVSFEELTFGHDAKFIDSLQENSRFLYYNKKFDIASTLNKA	300		
Dd	241	RVFKVNTNAYEMSGLEVSFEELTFGHDAKFIDSLQENSRFLYYNKKFDIASTLNKA	300		
Qy	301	KSIVGTTASLOYMKNVKPKYLLSDETSKGFSDVKLKFDPKLYKMLTEIYTEDNFKVFKV	360		
Dd	301	KSIVGTTASLOYMKNVKPKYLLSDETSKGFSDVKLKFDPKLYKMLTEIYTEDNFKVFKV	360		
Qy	361	LNKRTYLNFDKAVFKINIVPKVNTIYDGNLRLNTNLAANPNGONTENNMTKLKNET	420		
Dd	361	LNKRTYLNFDKAVFKINIVPKVNTIYDGNLRLNTNLAANPNGONTENNMTKLKNET	420		
Qy	421	GLFEFYKLLCVRGIIITKTSKLDGYNKALNDLCIKVNNMDLFFSPSEDNFTDNLKGE	480		
Dd	421	GLFEFYKLLCVRGIIITKTSKLDGYNKALNDLCIKVNNMDLFFSPSEDNFTDNLKGE	480		
Qy	481	ITSDTNTIEAENISLDLIQOYYITFNFDNPNENISNTIENSSDIIGOLELMPNTERP	540		
Dd	481	ITSDTNTIEAENISLDLIQOYYITFNFDNPNENISNTIENSSDIIGOLELMPNTERP	540		
Qy	541	KKYELDKYTMHYLRAQEFEGKGRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA	600		
Dd	541	KKYELDKYTMHYLRAQEFEGKGRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA	600		
Qy	601	AMELQWELVYDFTDETSEVSTTDKADIITIIPIYGPALNIGNMLYKDDFVQALLFSG	660		
Dd	601	AMELQWELVYDFTDETSEVSTTDKADIITIIPIYGPALNIGNMLYKDDFVQALLFSG	660		
Qy	661	AVILLEPIEAI	PVLGTFALVSYIANKVLTVOTIDNLSKRNEKWDVYKIVTNWLAK	720	


```
Qy 241 RVFKNVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLTKA 300
Db 241 RVFKNVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLTKA 300
Qy 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDFKLYKMLTEIYEDNFVKPKV 360
Db 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDFKLYKMLTEIYEDNFVKPKV 360
Qy 361 LNRKTYLNFKAQVFKINIVPKVNTIYDGFNLRTNLAAFNPGQNTINNMFNFKLNFT 420
Db 361 LNRKTYLNFKAQVFKINIVPKVNTIYDGFNLRTNLAAFNPGQNTINNMFNFKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
Db 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
Qy 481 ITSDTNEAAEENISLDLIQOYVLTFFNDEPENISLENSSDIIGOLELMPNTERFNG 540
Db 481 ITSDTNEAAEENISLDLIQOYVLTFFNDEPENISLENSSDIIGOLELMPNTERFNG 540
Qy 541 KYELDKYTFPHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Db 541 KYELDKYTFPHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
Qy 661 AVILLEFIPEIAPVLGTFFALVSIAKVLTVQIDNALSKRNEKWDVYKIVTNWLAK 720
Db 661 AVILLEFIPEIAPVLGTFFALVSIAKVLTVQIDNALSKRNEKWDVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKALENOAEATKAIINYQVNOYTEEEKNNINIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKALENOAEATKAIINYQVNOYTEEEKNNINIDDLSSKLNESINKA 780
Qy 781 MININKEFNQCSVSYLMSMIPYGVKRLDQFASLKDALLKYIIDNRTGLIGQVDRLLK 840
Db 781 MININKEFNQCSVSYLMSMIPYGVKRLDQFASLKDALLKYIIDNRTGLIGQVDRLLK 840
Qy 841 VNTLSTDIIPQLSKYVDNQRLLTFTTEYIK 871
Db 841 VNTLSTDIIPQLSKYVDNQRLLTFTTEYIK 871
```

RESULT 7

```
US-09-255-829-26
; Sequence 26, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
```

```
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-255-829-26
```

```
Query Match 66.1%; Score 4507; DB 4; Length 871;
Best Local Similarity 99.8%; Pred. No. 6.6e-295;
Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVNFQVNFYKDPVNGVDIAIYIKIPNAGOMQPVKAPKIHNNKIWVPERDFTNPEEGDLN 60
Db 1 MQFVNFQVNFYKDPVNGVDIAIYIKIPNAGOMQPVKAPKIHNNKIWVPERDFTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYSDYSLSTDNEKDYLGKVTGLFERIYSTDLGRMLLTISVIRGIPFWGG 120
Db 61 PPPEAKQVPVSYSDYSLSTDNEKDYLGKVTGLFERIYSTDLGRMLLTISVIRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIOPECKSGFHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIOPECKSGFHEVLNLTNGY 180
Qy 181 GSTQYIRPSPDPTFGPEESLEVDNPLLGAGKATDPAVTLAHLIIYAGHRLYGIANPN 240
Db 181 GSTQYIRPSPDPTFGPEESLEVDNPLLGAGKATDPAVTLAHLIIYAGHRLYGIANPN 240
Qy 241 RVFKNVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLTKA 300
Db 241 RVFKNVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLTKA 300
Qy 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDFKLYKMLTEIYEDNFVKPKV 360
Db 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDFKLYKMLTEIYEDNFVKPKV 360
Qy 361 LNRKTYLNFKAQVFKINIVPKVNTIYDGFNLRTNLAAFNPGQNTINNMFNFKLNFT 420
Db 361 LNRKTYLNFKAQVFKINIVPKVNTIYDGFNLRTNLAAFNPGQNTINNMFNFKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
Db 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
Qy 481 ITSDTNEAAEENISLDLIQOYVLTFFNDEPENISLENSSDIIGOLELMPNTERFNG 540
Db 481 ITSDTNEAAEENISLDLIQOYVLTFFNDEPENISLENSSDIIGOLELMPNTERFNG 540
Qy 541 KYELDKYTFPHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Db 541 KYELDKYTFPHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
Qy 661 AVILLEFIPEIAPVLGTFFALVSIAKVLTVQIDNALSKRNEKWDVYKIVTNWLAK 720
Db 661 AVILLEFIPEIAPVLGTFFALVSIAKVLTVQIDNALSKRNEKWDVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKALENOAEATKAIINYQVNOYTEEEKNNINIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKALENOAEATKAIINYQVNOYTEEEKNNINIDDLSSKLNESINKA 780
```

Db 721 VNTQIDLRKMKKALENQAEBATKAIINYQYNTVEBEKNNINFNIDDLGSKLNESINKA 780
Qy 781 MININKFLNQCYSVLYMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Db 781 MININKFLNQCYSVLYMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Qy 841 VNNLTSTDIPFQLSKYVDNORLLSTFTTEYIK 871
Db 841 VNNLTSTDIPFQLSKYVDNORLLSTFTTEYIK 871

RESULT 8
US-09-255-829-18
; Sequence 18, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1013 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-18

Query Match 66.0%; Score 4501; DB 4; Length 1013;
Best Local Similarity 95.5%; Pred. No. 2.1e-294;
Matches 877; Conservative 3; Mismatches 12; Indels 26; Gaps 2;

Qy 1 MPFVNKQNPYKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIPIERDFTFNPEEGDLN 60
Db 1 MQFVNKQNPYKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIPIERDFTFNPEEGDLN 60
Qy 61 PPPKAKQVPVSYDYSTYLTSTONKDYKLVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPPKAKQVPVSYDYSTYLTSTONKDYKLVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPGSVRSBELNVIIGPSADIIOFECKSPGHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPGSVRSBELNVIIGPSADIIOFECKSPGHEVLNLTNGY 180

Qy 181 GSTQYIRFSPDFTFGFEESELEVDNINPLLGAGKATDPAVTLAHELIYAGHRLYGIAPN 240
Db 181 GSTQYIRFSPDFTFGFEESELEVDNINPLLGAGKATDPAVTLAHELIYAGHRLYGIAPN 240
Qy 241 RVKQVNTNAYEMSGLEVSFEELRTFGHDAKFTDSLOENEFRLLYYNKKDIASTLNKA 300
Db 241 RVKQVNTNAYEMSGLEVSFEELRTFGHDAKFTDSLOENEFRLLYYNKKDIASTLNKA 300
Qy 301 KSIIVGTTASLOYKMNVPFKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYYEDNFVKPFKV 360
Db 301 KSIIVGTTASLOYKMNVPFKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYYEDNFVKPFKV 360
Qy 361 LNRKTYLNFDAVKPKINIVPKVNYIYDGFNLNLTNLAANFGONTINNNFTKLKQFT 420
Db 361 LNRKTYLNFDAVKPKINIVPKVNYIYDGFNLNLTNLAANFGONTINNNFTKLKQFT 420
Qy 421 GLPEFYKLLCVRGIIITSKTSKLDGYNK-----ALNDLCIKVNNWDLFPSEEDNPTN 473
Db 421 GLPEFYKLLCVRGIIITSKTSKLDGYNKIEGRCDGALNDLCIKVNNWDLFPSEEDNPTN 480
Qy 474 DLNKGEEITSDTNIEAAENISLDLIQOYYLTFTNFNEPENISLENSSDIIGOLELMPN 533
Db 481 DLNKGEEITSDTNIEAAENISLDLIQOYYLTFTNFNEPENISLENSSDIIGOLELMPN 540
Qy 534 IERPPNGKCYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKK 593
Db 541 IERPPNGKCYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKK 600
Qy 594 VNKA TEAAMEFLGWVEQLVYDFTDETSEVSTTDKTDITIIPIYIGPALNIGNMLYKDDFV 653
Db 601 VNKA TEAAMEFLGWVEQLVYDFTDETSEVSTTDKTDITIIPIYIGPALNIGNMLYKDDFV 660
Qy 654 GALIFSGAVILLEFIPETALPVLGTFAVSYIAKVKLTQVTDNALSKRNEKWDVYKVI 713
Db 661 GALIFSGAVILLEFIPETALPVLGTFAVSYIAKVKLTQVTDNALSKRNEKWDVYKVI 720
Qy 714 VTNWLAKVNTQIDLRKMKKALENQAEBATKAIINYQYNTVEBEKNNINFNIDDLGSKLN 773
Db 721 VTNWLAKVNTQIDLRKMKKALENQAEBATKAIINYQYNTVEBEKNNINFNIDDLGSKLN 780
Qy 774 NESINKAMINIKFLNQCYSVLYMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQ 833
Db 781 NESINKAMINIKFLNQCYSVLYMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQ 840
Qy 834 VDRLLKDKVNTLTSTDIPFQLSKYVDNORLLSTFTTEYIK-----NII 874
Db 841 VDRLLKDKVNTLTSTDIPFQLSKYVDNORLLSTFTTEYIKSGLSNPGAAHYAQHDEAVDNKF 900

875 NTSILNURYESNHLIDLS 892
901 NKEQONAFYEILHLPNLN 918

RESULT 9
US-09-255-829-6
; Sequence 6, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible


```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 873 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-255-829-6

Query Match 65.9%; Score 4495; DB 4; Length 873;
Best Local Similarity 99.3%; Pred. No. 4.2e-294;
Matches 867; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPFVKQFNKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Db 3 MEFVNKQFNKDPVNGVDIAIYIKIPKYGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 62
Qy 61 PPEAKQVPVSYDSTYLSSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTIVRGIPFWGG 120
Db 63 PPEAKQVPVSYDSTYLSSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTIVRGIPFWGG 122
Qy 121 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY 180
Db 123 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY 182
Qy 181 GSTQYIRFSDPTFGPESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db 183 GSTQYIRFSDPTFGPESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 242
Qy 241 RVKVNNTAYEMSGLEVSFEELTFGCHDAKFTDSIQENEFRLYYNKKFDIATSLNKA 300
Db 243 RVKVNNTAYEMSGLEVSFEELTFGCHDAKFTDSIQENEFRLYYNKKFDIATSLNKA 302
Qy 301 KSIIVGTTASLQYMKNVFKEKYLLEDSTSGKFSVDKLFKXLYKMLBIYTEDNFVKPFKV 360
Db 303 KSIIVGTTASLQYMKNVFKEKYLLEDSTSGKFSVDKLFKXLYKMLBIYTEDNFVKPFKV 362
Qy 361 LNRKTYLNFKAQVKNIVPKVNTIYDGNFLRNLTNLAANFNGONTINNNTFKLNFT 420
Db 363 LNRKTYLNFKAQVKNIVPKVNTIYDGNFLRNLTNLAANFNGONTINNNTFKLNFT 422
Qy 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKYNNNDLPSPEDNFTNDLNKGE 480
Db 423 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKYNNNDLPSPEDNFTNDLNKGE 482
Qy 481 ITSNTIEAAEENISLDLIQOYLYTFNFDNEPENISLENSSDIIGQELMNPNIERPNG 540
Db 483 ITSNTIEAAEENISLDLIQOYLYTFNFDNEPENISLENSSDIIGQELMNPNIERPNG 542
Qy 541 KKYELDKYTFHYLRAQEFEGHKSRIALTNSVNEALLNPGRVYTFPSSDYVKKVKNKATEA 600
Db 543 KKYELDKYTFHYLRAQEFEGHKSRIALTNSVNEALLNPGRVYTFPSSDYVKKVKNKATEA 602
Qy 601 AMFLGWVQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 660
Db 603 AMFLGWVQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 662

661 AVILLEPIEPIAIPVLGTFFALVSVIANKVLTVQIDNALSKRNEKQWDEVYKIVTNMLAK 720
663 AVILLEPIEPIAIPVLGTFFALVSVIANKVLTVQIDNALSKRNEKQWDEVYKIVTNMLAK 722
721 VNTQIDILIRKKMEALENQAEATKAIINYQVQYVTEEEKNNINENIDDLSSKLNESINKA 780
723 VNTQIDILIRKKMEALENQAEATKAIINYQVQYVTEEEKNNINENIDDLSSKLNESINKA 782
781 MININKFLNQCYSVYLNMSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRCLKDK 840
783 MININKFLNQCYSVYLNMSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRCLKDK 842
841 VNNLTSTDIPFQLSKYVDNQRLLSTFTTEYIK 871
843 VNNLTSTDIPFQLSKYVDNQRLLSTFTTEYIK 873

RESULT 10
US-09-255-829-10
; Sequence 10, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 875 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-255-829-10

Query Match 65.9%; Score 4495; DB 4; Length 875;
Best Local Similarity 99.3%; Pred. No. 4.2e-294;
Matches 869; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

Qy 1 MPFVKQFNKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Db 1 MQFVNKQFNKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYDSTYLSSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTIVRGIPFWGG 120
```


Db 61 PPEAKQPVSYDSTVLTSTNEKDNVLTGVTGKLFERIYSTDLGRMLLTIVRGIPFWGG 120
Qy 121 STIDTELKVDITNCINVIQDGSYRSEELNVLIGPSADIIQPECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVDITNCINVIQDGSYRSEELNVLIGPSADIIQPECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSDFTFGPESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db 181 GSTQYIRFSDFTFGPESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKPDIASLTANKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKPDIASLTANKA 300
Qy 301 KSVIGTTASIQYMKNVFKEKYLLEDTSFGKFSVDKLFKDYKMLTEIYEDNFVKFVK 360
Db 301 KSVIGTTASIQYMKNVFKEKYLLEDTSFGKFSVDKLFKDYKMLTEIYEDNFVKFVK 360
Qy 361 LNRKTYLNFDAKVFKNIVPKVNYTIYDGFNLRTNLAANFNGQNTINNMFVKLKNFT 420
Db 361 LNRKTYLNFDAKVFKNIVPKVNYTIYDGFNLRTNLAANFNGQNTINNMFVKLKNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKTSKLDKGNK----ALNDLCIKVNNWDLFPSPSEDNFTNDLN 476
Db 421 GLFEFYKLLCVRGIIITSKTSKLDKGNKADGALNDLCIKVNNWDLFPSPSEDNFTNDLN 480
Qy 477 KGBEITSDTNEAAEENISLDLIQOYLTNFNDNEPENISLENSSDIIGOLELMPNIE 536
Db 481 KGBEITSDTNEAAEENISLDLIQOYLTNFNDNEPENISLENSSDIIGOLELMPNIE 540
Qy 537 PPNKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKN 596
Db 541 PPNKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKN 600
Qy 597 ATEAAMFLGWVQVYDFTDTSVSTTKIADITIIIPYGPALNTGNMLYKDDFVGAL 656
Db 601 ATEAAMFLGWVQVYDFTDTSVSTTKIADITIIIPYGPALNTGNMLYKDDFVGAL 660
Qy 657 IFSGAVILLEPIEPIALPVLTGTFALVSIAKVLTVOTIDNALSKEKDEYVYKIVTN 716
Db 661 IFSGAVILLEPIEPIALPVLTGTFALVSIAKVLTVOTIDNALSKEKDEYVYKIVTN 720
Qy 717 WLAKVNTQIDILIRKMKALENAEATKAIINYQYNTYEEKNINFNIDDLSSKLNES 776
Db 721 WLAKVNTQIDILIRKMKALENAEATKAIINYQYNTYEEKNINFNIDDLSSKLNES 780
Qy 777 INKAMINIKFLNOCQSVSYLMSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGQVDR 836
Db 781 INKAMINIKFLNOCQSVSYLMSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGQVDR 840
Qy 837 LKDKVNTLSTDIPFQLSKYVDNORLLSTFTFYIK 871
Db 841 LKDKVNTLSTDIPFQLSKYVDNORLLSTFTFYIK 875

RESULT 11

US-09-255-829-4
; Sequence 4, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
CURRENT APPLICATION DATA:
FILING DATE: 23-FEB-1999
APPLICATION NUMBER: US/09/255,829
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/02273
FILING DATE: 22-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,893
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1581.0130002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-255-829-4

Query Match 65.9%; Score 4495; DB 4; Length 894;

Best Local Similarity 99.5%; Pred. No. 4.4e-294;

Matches 867; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPFNQPNFYKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIPIERDFTTNPEDGLN 60
Db 24 MEFVNQPNFYKDPVNGVDIAYIKIPKYGQMPVKAFAKIHKNKIWIPIERDFTTNPEDGLN 83
Qy 61 PPEAKQPVSYDSTVLTSTNEKDNVLTGVTGKLFERIYSTDLGRMLLTIVRGIPFWGG 120
Db 84 PPEAKQPVSYDSTVLTSTNEKDNVLTGVTGKLFERIYSTDLGRMLLTIVRGIPFWGG 143
Qy 121 STIDTELKVDITNCINVIQDGSYRSEELNVLIGPSADIIQPECKSFGEHVLNLTNGY 180
Db 144 STIDTELKVDITNCINVIQDGSYRSEELNVLIGPSADIIQPECKSFGEHVLNLTNGY 203
Qy 181 GSTQYIRFSDFTFGPESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db 204 GSTQYIRFSDFTFGPESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 263
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKPDIASLTANKA 300
Db 264 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKPDIASLTANKA 323
Qy 301 KSVIGTTASIQYMKNVFKEKYLLEDTSFGKFSVDKLFKDYKMLTEIYEDNFVKFVK 360
Db 324 KSVIGTTASIQYMKNVFKEKYLLEDTSFGKFSVDKLFKDYKMLTEIYEDNFVKFVK 383
Qy 361 LNRKTYLNFDAKVFKNIVPKVNYTIYDGFNLRTNLAANFNGQNTINNMFVKLKNFT 420
Db 384 LNRKTYLNFDAKVFKNIVPKVNYTIYDGFNLRTNLAANFNGQNTINNMFVKLKNFT 443
Qy 421 GLFEFYKLLCVRGIIITSKTSKLDKGNKADGALNDLCIKVNNWDLFPSPSEDNFTNDLNK 480
Db 444 GLFEFYKLLCVRGIIITSKTSKLDKGNKADGALNDLCIKVNNWDLFPSPSEDNFTNDLNK 503
Qy 481 ITSDTNEAAEENISLDLIQOYLTNFNDNEPENISLENSSDIIGOLELMPNIEFPNG 540
Db 504 ITSDTNEAAEENISLDLIQOYLTNFNDNEPENISLENSSDIIGOLELMPNIEFPNG 563
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Db 564 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 623


```
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/255,829
FILING DATE: 23-FEB-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/02273
FILING DATE: 22-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,893
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1581.0130002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 953 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-255-829-14

Query Match 65.9%; Score 4494.5; DB 4; Length 953;
Best Local Similarity 98.0%; Pred. No. 5.2e-294;
Matches 869; Conservative 2; Mismatches 1; Indels 7; Gaps 1;

Qy 1 MPFVNKQPNKDPVNGVDIAIKIPNAGQMPVKAQKIHKKIWIPIERDPTNPEEGDLN 60
Db 1 MQFVNKQPNKDPVNGVDIAIKIPNAGQMPVKAQKIHKKIWIPIERDPTNPEEGDLN 60

Qy 61 PPEAKQVPVSYDSTYLTSTNEXNDYLGKTKLPERIYSTDLGRMLTSTVRCIPFWGG 120
Db 61 PPEAKQVPVSYDSTYLTSTNEXNDYLGKTKLPERIYSTDLGRMLTSTVRCIPFWGG 120

Qy 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQFCKSGFHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQFCKSGFHEVLNLTNGY 180

Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLIGAGKFAITDPAVTLAHLIYAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGFESLEVDNPLIGAGKFAITDPAVTLAHLIYAGHRLYGIAPNP 240

Qy 241 RVFKVNTNAYEMSGLEVSFEELATFGHDAKFDLSQENEFRLYYNKKDIASTLNKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELATFGHDAKFDLSQENEFRLYYNKKDIASTLNKA 300

Qy 301 KSIIVGTTASLOYMKNVKFKYLLSEDTSGKFSVDKLFKDFKLYKMLTEIYEDNFVKFKV 360
Db 301 KSIIVGTTASLOYMKNVKFKYLLSEDTSGKFSVDKLFKDFKLYKMLTEIYEDNFVKFKV 360

Qy 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFGQNTENNMFYKLNFT 420
Db 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFGQNTENNMFYKLNFT 420

Qy 421 GLFSFYKLLCVRGIIITSKTSLDGYNK-----ALNDLCIKVNNWDLFSPSEDFTN 473
Db 421 GLFSFYKLLCVRGIIITSKTSLDGYNKIEGRCDGALNDLCIKVNNWDLFSPSEDFTN 480

Qy 474 DLNKGEEITSDTNEAAEENISLDLIQOYLYLTFFNDEPENISLENLSSDIIGOLEMPN 533
Db 481 DLNKGEEITSDTNEAAEENISLDLIQOYLYLTFFNDEPENISLENLSSDIIGOLEMPN 540

Qy 534 IERFPNGKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKK 593
Db 534 IERFPNGKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKK 600

Qy 594 VNKATEAMFLGWVEQLVYDFTDSTSEVSTTKDIADITIIPIYGPALNIGNMLYKDDFV 653
Db 601 VNKATEAMFLGWVEQLVYDFTDSTSEVSTTKDIADITIIPIYGPALNIGNMLYKDDFV 660

Qy 654 GALIFSGAVILLEPIEIPALVLTGTFALSVYANKVLTQVOTIDNALSKRNEKMDVYKYI 713
Db 661 GALIFSGAVILLEPIEIPALVLTGTFALSVYANKVLTQVOTIDNALSKRNEKMDVYKYI 720

Qy 714 VTNWLAKVNTQIDILIRKMKKEALENQAETKAIINYOYNTTEBEKKNINFNIDDLSSKL 773
Db 721 VTNWLAKVNTQIDILIRKMKKEALENQAETKAIINYOYNTTEBEKKNINFNIDDLSSKL 780

Qy 774 NESINKAMINIKFLNOCVSYLMNSMIPYGVKRELFDFASLKDALLKYYIDNRGTLIGQ 833
Db 781 NESINKAMINIKFLNOCVSYLMNSMIPYGVKRELFDFASLKDALLKYYIDNRGTLIGQ 840

Qy 834 VDLRKQKVNNTLTSTDIIPFQLSKYVDNQRLLSTFTFYIKN 872
Db 841 VDLRKQKVNNTLTSTDIIPFQLSKYVDNQRLLSTFTFYIKS 879

RESULT 14
US-09-255-829-12
; Sequence 12, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-255-829-12

Query Match 65.9%; Score 4493.5; DB 4; Length 878;
Best Local Similarity 99.0%; Pred. No. 5.4e-294;
Matches 869; Conservative 1; Mismatches 1; Indels 7; Gaps 1;
```

```
QY 1 MPFVNQFNKDPVNGVDIAYIKIPNAGQMPKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
|
|
|
Db 1 MQFVNQFNKDPVNGVDIAYIKIPNAGQMPKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
|
|
|
QY 61 PPPEAKQVPVSYDSTYSLTDNEKDNKYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
|
|
|
Db 61 PPPEAKQVPVSYDSTYSLTDNEKDNKYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
|
|
|
QY 121 STIDTELKVIDTNCINVIQDGSVRSEELNVIIGPSADIIQPECKSFGEHVLNLTNGY 180
|
|
|
Db 121 STIDTELKVIDTNCINVIQDGSVRSEELNVIIGPSADIIQPECKSFGEHVLNLTNGY 180
|
|
|
QY 181 GSTQYIRFSPDFTFGFEESLEVDTPNLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
|
|
|
Db 181 GSTQYIRFSPDFTFGFEESLEVDTPNLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
|
|
|
QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKKFDIATSLNKA 300
|
|
|
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKKFDIATSLNKA 300
|
|
|
QY 301 KSIIVGTTASIQYMKNVFKEKYLLEDTSKGFSVDKLFKDKLYKMLTEIYTEDNPFVKFKV 360
|
|
|
Db 301 KSIIVGTTASIQYMKNVFKEKYLLEDTSKGFSVDKLFKDKLYKMLTEIYTEDNPFVKFKV 360
|
|
|
QY 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRLNTNLAANFNQNTNINNNFTKLKNT 420
|
|
|
Db 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRLNTNLAANFNQNTNINNNFTKLKNT 420
|
|
|
QY 421 GLPEFYKLLCVRGIIITSKTSLDKGYNK-----ALNDLCIKVNNWDLFFSPSEDNFTN 473
|
|
|
Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKIEGRCDGALNDLCIKVNNWDLFFSPSEDNFTN 480
|
|
|
QY 474 DLNKGSEITSDTNEAAEENISLDLIQOYVLTFTNFDNEPENISLENSSDIIIGOLELMPN 533
|
|
|
Db 481 DLNKGSEITSDTNEAAEENISLDLIQOYVLTFTNFDNEPENISLENSSDIIIGOLELMPN 540
|
|
|
QY 534 IERPNKGKVELDKYTMFHYLRAQEFEGHKSRIALTNSVNEALLNPSRVYTFSSDYVKK 593
|
|
|
Db 541 IERPNKGKVELDKYTMFHYLRAQEFEGHKSRIALTNSVNEALLNPSRVYTFSSDYVKK 600
|
|
|
QY 594 VNKAATEAMFLGWVEQLVYDFTDSTSEVSTDKIADITIIPIYIGPALNIGNMLYKDDFV 653
|
|
|
Db 601 VNKAATEAMFLGWVEQLVYDFTDSTSEVSTDKIADITIIPIYIGPALNIGNMLYKDDFV 660
|
|
|
QY 654 GALIFSGAVILLEFIPETAPVLGTFALVSYIANKVLTQVTDNALSKRNEKWDVYKII 713
|
|
|
Db 661 GALIFSGAVILLEFIPETAPVLGTFALVSYIANKVLTQVTDNALSKRNEKWDVYKII 720
|
|
|
QY 714 VTNNLAKVNTQIDILIRKQKKEALENQAETKAIINYOYQVNTYEEKKNINFNIDDLSSKL 773
|
|
|
Db 721 VTNNLAKVNTQIDILIRKQKKEALENQAETKAIINYOYQVNTYEEKKNINFNIDDLSSKL 780
|
|
|
QY 774 NESINKAMINIKFNQCSVSYLWMSIPIYGVKRELEDFDASLKDALLKYIYDNRGTIIGQ 833
|
|
|
Db 781 NESINKAMINIKFNQCSVSYLWMSIPIYGVKRELEDFDASLKDALLKYIYDNRGTIIGQ 840
|
|
|
QY 834 VDLRKDKVNTLSTDDIPQLSKYVDNORLLSTFTEYIK 871
|
|
|
Db 841 VDLRKDKVNTLSTDDIPQLSKYVDNORLLSTFTEYIK 878
|
|
|
```

RESULT 15

```
US-09-288-326A-7
; Sequence 7, Application US/09288326A
; Patent No. 6776990
; GENERAL INFORMATION:
; APPLICANT: Steward, Lance E.
; APPLICANT: Aoki, K. Roger
; APPLICANT: Sachs, George
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Treatment of Pancreatitis
; FILE REFERENCE: 17282
```

```
; CURRENT APPLICATION NUMBER: US/09/288,326A
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-288-326A-7

Query Match 34.4%; Score 2348; DB 4; Length 448;
Best Local Similarity 99.8%; Pred. No. 5.2e-150; Indels 0; Gaps 0;
Matches 447; Conservative 1; Mismatches 0;

QY 1 MPFVNQFNKDPVNGVDIAYIKIPNAGQMPKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
|
|
|
Db 1 MPFVNQFNKDPVNGVDIAYIKIPNAGQMPKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
|
|
|
QY 61 PPPEAKQVPVSYDSTYSLTDNEKDNKYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
|
|
|
Db 61 PPPEAKQVPVSYDSTYSLTDNEKDNKYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
|
|
|
QY 121 STIDTELKVIDTNCINVIQDGSVRSEELNVIIGPSADIIQPECKSFGEHVLNLTNGY 180
|
|
|
Db 121 STIDTELKVIDTNCINVIQDGSVRSEELNVIIGPSADIIQPECKSFGEHVLNLTNGY 180
|
|
|
QY 181 GSTQYIRFSPDFTFGFEESLEVDTPNLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
|
|
|
Db 181 GSTQYIRFSPDFTFGFEESLEVDTPNLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
|
|
|
QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKKFDIATSLNKA 300
|
|
|
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKKFDIATSLNKA 300
|
|
|
QY 301 KSIIVGTTASIQYMKNVFKEKYLLEDTSKGFSVDKLFKDKLYKMLTEIYTEDNPFVKFKV 360
|
|
|
Db 301 KSIIVGTTASIQYMKNVFKEKYLLEDTSKGFSVDKLFKDKLYKMLTEIYTEDNPFVKFKV 360
|
|
|
QY 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRLNTNLAANFNQNTNINNNFTKLKNT 420
|
|
|
Db 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRLNTNLAANFNQNTNINNNFTKLKNT 420
|
|
|
QY 421 GLPEFYKLLCVRGIIITSKTSLDKGYNK 448
|
|
|
Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYNK 448
|
|
|

Search completed: January 31, 2005, 14:09:56
Job time : 48 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:04 ; Search time 174 Seconds
(without alignments)
2703.441 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFVNQFNKDPVNGVDIA.....EFIPVDGNGERPLHHHHH 1302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6804	99.8	1306	15	US-10-205-516-16
2	6768	99.2	1296	15	US-10-452-024-160
3	6768	99.2	1296	15	US-10-205-516-2
4	6764	99.2	1296	15	US-10-452-024-159
5	6764	99.2	1296	17	US-10-821-669-1
6	6759	99.1	1295	9	US-09-726-949A-1
7	6759	99.1	1295	15	US-10-452-024-157
8	6756	99.0	1296	14	US-10-272-898-28
9	6756	99.0	1296	14	US-10-354-774-28
10	6756	99.0	1296	14	US-10-371-012-28
11	6756	99.0	1296	15	US-10-452-024-158
12	6756	99.0	1296	16	US-10-729-122-28
13	6756	99.0	1296	16	US-10-729-039-28

14	6756	99.0	1296	17	US-10-729-527-28	Sequence 28, Appl
15	6756	99.0	1296	17	US-10-727-898-28	Sequence 28, Appl
16	6756	99.0	1296	17	US-10-728-696-28	Sequence 28, Appl
17	6158	90.3	1296	15	US-10-452-024-1	Sequence 1, Appl
18	6158	90.3	1296	15	US-10-452-024-130	Sequence 130, Appl
19	4515	66.2	871	9	US-09-255-829-8	Sequence 8, Appl
20	4515	66.2	871	14	US-10-241-596-8	Sequence 8, Appl
21	4512	66.1	871	14	US-10-241-596-155	Sequence 155, App
22	4512	66.1	873	14	US-10-241-596-163	Sequence 163, App
23	4512	66.1	873	14	US-10-241-596-165	Sequence 165, App
24	4509	66.1	873	14	US-10-241-596-169	Sequence 169, App
25	4507	66.1	871	9	US-09-255-829-2	Sequence 2, Appl
26	4507	66.1	871	9	US-09-255-829-26	Sequence 26, Appl
27	4507	66.1	871	14	US-10-241-596-2	Sequence 2, Appl
28	4507	66.1	871	14	US-10-241-596-26	Sequence 26, Appl
29	4507	66.1	871	14	US-10-241-596-153	Sequence 153, App
30	4507	66.1	871	15	US-10-452-024-154	Sequence 154, App
31	4507	66.1	873	14	US-10-241-596-149	Sequence 149, App
32	4507	66.1	873	14	US-10-241-596-151	Sequence 151, App
33	4507	66.1	879	14	US-10-241-596-159	Sequence 159, App
34	4507	66.1	1127	14	US-10-241-596-40	Sequence 40, Appl
35	4507	66.1	1129	14	US-10-241-596-42	Sequence 42, Appl
36	4504	66.0	873	14	US-10-241-596-167	Sequence 167, App
37	4501	66.0	1013	9	US-09-255-829-18	Sequence 18, Appl
38	4501	66.0	1013	14	US-10-241-596-18	Sequence 18, Appl
39	4495.5	65.9	1130	14	US-10-241-596-44	Sequence 44, Appl
40	4495.5	65.9	1130	14	US-10-241-596-139	Sequence 139, App
41	4495.5	65.9	1132	14	US-10-241-596-46	Sequence 46, Appl
42	4495	65.9	873	9	US-09-255-829-6	Sequence 6, Appl
43	4495	65.9	873	14	US-10-241-596-6	Sequence 6, Appl
44	4495	65.9	875	9	US-09-255-829-10	Sequence 10, Appl
45	4495	65.9	875	14	US-10-241-596-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-10-205-516-16
; Sequence 16, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: jzbt.xl
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1306
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-16

Query Match 99.8%; Score 6804; DB 15; Length 1306;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1301; Conservative 1; Mismatches 0; Indels 4; Gaps 1;

Qy	1	MPFVNQFNKDPVNGVDIAIYIKIPNAGQMPKAFKHKNKIWIIPERDFTTPEEGDLN	60
Db	1	MPFVNQFNKDPVNGVDIAIYIKIPNAGQMPKAFKHKNKIWIIPERDFTTPEEGDLN	60
Qy	61	PPPEAKQVPVSYDSTYLSITDNEKDNLYKGVTKLFEIYSTDIGRMLLTSIVRGIPWGG	120
Db	61	PPPEAKQVPVSYDSTYLSITDNEKDNLYKGVTKLFEIYSTDIGRMLLTSIVRGIPWGG	120
Qy	121	STIDTELKVIDTNCINVIQDGSYRSEELNLVIIGPSADIIOQECKSFGEHVLNLTNGY	180
Db	121	STIDTELKVIDTNCINVIQDGSYRSEELNLVIIGPSADIIOQECKSFGEHVLNLTNGY	180
Qy	181	GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKATDPAVTLAHELIYAGHRLYGIATNP	240

181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIHAHRLYGIAPNP 240
241 RVKQNTNAYEMSGLEVSPEELRTFGCHDAKFTDSIQENEFRLYYNKKFDIATSLNKA 300
241 RVKQNTNAYEMSGLEVSPEELRTFGCHDAKFTDSIQENEFRLYYNKKFDIATSLNKA 300
301 KSIIVGTTASLOYMNKVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
301 KSIIVGTTASLOYMNKVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRTNLAAFPNGQNTENNMFVKLNFT 420
361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRTNLAAFPNGQNTENNMFVKLNFT 420
421 GLPEFYKLLCVRGIIITSTKSLDGKYNK---ALNDLCIKVNNWDLFPSPSEDNFTNDLN 476
421 GLPEFYKLLCVRGIIITSTKSLDGKYNK---ALNDLCIKVNNWDLFPSPSEDNFTNDLN 480
477 KGEIITSDTNEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIE 536
481 KGEIITSDTNEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIE 540
537 FPNKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALINPNSRVVTFPSSDYVKKVK 596
541 FPNKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALINPNSRVVTFPSSDYVKKVK 600
597 ATEAMFLGWVEQLVDFDTSETSVSTTKIADITIIIPYIGPALNIGNMLYKDDFVGL 656
601 ATEAMFLGWVEQLVDFDTSETSVSTTKIADITIIIPYIGPALNIGNMLYKDDFVGL 660
657 IFSGAVILLETPIEIALPVLTGTFALVSIAKVLTVQTDNALSKRNEKDEVKYIWTN 716
661 IFSGAVILLETPIEIALPVLTGTFALVSIAKVLTVQTDNALSKRNEKDEVKYIWTN 720
717 WLAKVNTQIDILIRKMEALENQAETAKAIINYQNYQVTEBEKNINFNIDDLSSKLNES 776
721 WLAKVNTQIDILIRKMEALENQAETAKAIINYQNYQVTEBEKNINFNIDDLSSKLNES 780
777 INKAMININFLNCSVSYLMNSMIPYGVKLEDFDASLKDALIKYIDNRGTILGOVDR 836
781 INKAMININFLNCSVSYLMNSMIPYGVKLEDFDASLKDALIKYIDNRGTILGOVDR 840
837 LKDKVNTLSTDIPFQLSKYVDNORLLSTFTEYIKNIINTSILNRYESNHLIDLSRYAS 896
841 LKDKVNTLSTDIPFQLSKYVDNORLLSTFTEYIKNIINTSILNRYESNHLIDLSRYAS 900
897 KINIGSKVNFDPIDKQIQLFNLESSKIEVLKNAIVNYSMYENFSTFWIRIPKFN 956
901 KINIGSKVNFDPIDKQIQLFNLESSKIEVLKNAIVNYSMYENFSTFWIRIPKFN 960
957 SLNNEYTIINCMENNNGKVSILNGEIIWTLQDTQEIQRVVRKYQMINISDYINRWIF 1016
961 SLNNEYTIINCMENNNGKVSILNGEIIWTLQDTQEIQRVVRKYQMINISDYINRWIF 1020
1017 VTIITNNLNNSKIYINGRLIDQKPISNLGNHASNIMFKDGCGRDTHRIWIKYFNLF 1076
1021 VTIITNNLNNSKIYINGRLIDQKPISNLGNHASNIMFKDGCGRDTHRIWIKYFNLF 1080
1077 KELNEKEIKOYDNQNSGILKDPWGDYLDKPYMMLNLYDPNKYVDVNVGIRGMYL 1136
1081 KELNEKEIKOYDNQNSGILKDPWGDYLDKPYMMLNLYDPNKYVDVNVGIRGMYL 1140
1137 KGPGRSVWNTNIIYLNSSIYGTKEPIIKYASGNKDNLRNDRVYINNVKKEIRLATN 1196
1141 KGPGRSVWNTNIIYLNSSIRGTKEPIIKYASGNKDNLRNDRVYINNVKKEIRLATN 1200
1197 ASQAGVEKILSALBIPDVGNLSQVVMKSKNDQGITNCKMNLQNDNGNDIGFIFGHQFN 1256
1201 ASQAGVEKILSALBIPDVGNLSQVVMKSKNDQGITNCKMNLQNDNGNDIGFIFGHQFN 1260
1257 NIAKLVASNWNQRQIERSRRTLGCSWEFIPVDDGWSGERPLHHHHH 1302

Db 1261 NIAKLVASNWNQRQIERSRRTLGCSWEFIPVDDGWSGERPLHHHHH 1306

RESULT 2

US-10-452-024-160
; Sequence 160, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 160
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; US-10-452-024-160

Query Match 99.2%; Score 6768; DB 15; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGQMPKAFKIHNKIWIIPERDFTNPEEGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGQMPKAFKIHNKIWIIPERDFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYSDYSTVLSTDNKDNVKGVTKLFBRIYSTDLGRMLTISIVRGIPFWGG 120
Db 61 PPEAKQVPVSYSDYSTVLSTDNKDNVKGVTKLFBRIYSTDLGRMLTISIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOFECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOFECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIHAHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIHAHRLYGIAPNP 240
Qy 241 RVKQNTNAYEMSGLEVSPEELRTFGCHDAKFTDSIQENEFRLYYNKKFDIATSLNKA 300
Db 241 RVKQNTNAYEMSGLEVSPEELRTFGCHDAKFTDSIQENEFRLYYNKKFDIATSLNKA 300
Qy 301 KSIIVGTTASLOYMNKVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
Db 301 KSIIVGTTASLOYMNKVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
Qy 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRTNLAAFPNGQNTENNMFVKLNFT 420
Db 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRTNLAAFPNGQNTENNMFVKLNFT 420
Qy 421 GLPEFYKLLCVRGIIITSTKSLDGKYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKEE 480
Db 421 GLPEFYKLLCVRGIIITSTKSLDGKYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKEE 480
Qy 481 ITSDTNIEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIEFPNG 540
Db 481 ITSDTNIEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIEFPNG 540
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALINPNSRVVTFPSSDYVKKVKATEA 600
Db 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALINPNSRVVTFPSSDYVKKVKATEA 600
Qy 601 AMFLGWVEQLVDFDTSETSVSTTKIADITIIIPYIGPALNIGNMLYKDDFVGLIFSG 660
Db 601 AMFLGWVEQLVDFDTSETSVSTTKIADITIIIPYIGPALNIGNMLYKDDFVGLIFSG 660


```

Qy 661 AVILLEPEIAIPVLGTFFALVSIAKVLTVQTDNALSKRNEKWDDEVYKIIVTNWLAK 720
Db 661 AVILLEPEIAIPVLGTFFALVSIAKVLTVQTDNALSKRNEKWDDEVYKIIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCSSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Db 781 MININKFLNOCSSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Qy 841 VNTLSTDIIPFQLSKYVDNQRLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Db 841 VNTLSTDIIPFQLSKYVDNQRLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Qy 901 GSKVNFDPIDKNQIQLFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKNQIQLFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTIINCMMNSGKVSILNYGEIITWLTQDTEIKQVVFVKYSQMINISDYINRWIFVTIT 1020
Db 961 EYTIINCMMNSGKVSILNYGEIITWLTQDTEIKQVVFVKYSQMINISDYINRWIFVTIT 1020
Qy 1021 NNRLNNSKIYINGRLIDQKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRLNNSKIYINGRLIDQKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKOLYDNQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Db 1081 EKEIKOLYDNQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Qy 1141 GSVMTNIIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
Db 1141 GSVMTNIIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGTNCKKNLQDNNGNDIGTGFQHFNNIAK 1260
Db 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGTNCKKNLQDNNGNDIGTGFQHFNNIAK 1260
Qy 1261 LVASNNYRQIERSSRTLGCSEWEPFVDDGWERPL 1296
Db 1261 LVASNNYRQIERSSRTLGCSEWEPFVDDGWERPL 1296

```

RESULT 3

US-10-205-516-2

```

; Sequence 2, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; FILE REFERENCE: JZbtxl
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1296
; TYPE: PR1
; ORGANISM: Clostridium botulinum
; US-10-205-516-2

```

```

Query Match 99.2%; Score 6768; DB 15; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps :0;

```

```

Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKHKNKIWIIPERDFTNPEEGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKHKNKIWIIPERDFTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYDSTYLDNEKDNLYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 120

```

```

Db 61 PPPEAKQVPVSYDSTYLDNEKDNLYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLVIGPSADIIQPECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLVIGPSADIIQPECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSDPDTFGPEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIANPN 240
Db 181 GSTQYIRFSDPDTFGPEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIANPN 240
Qy 241 RVFKVNTNAYENSGLEVSPEELRTFGGHDAKFIDSLQENEFRLYYNKKFDASTLNKA 300
Db 241 RVFKVNTNAYENSGLEVSPEELRTFGGHDAKFIDSLQENEFRLYYNKKFDASTLNKA 300
Qy 301 KSVIGTTASIQYMNKVFKEKYLLESSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
Db 301 KSVIGTTASIQYMNKVFKEKYLLESSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
Qy 361 LNRKTYLNPDKAVFKINIVPKVNYTYDGNLNRNTNLAANPNCQNTENNMTFKLNFT 420
Db 361 LNRKTYLNPDKAVFKINIVPKVNYTYDGNLNRNTNLAANPNCQNTENNMTFKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKTSKLDKYNKALNDLCIKVNNWDLFPSPSEDFTNDLNKGE 480
Db 421 GLFEFYKLLCVRGIIITSKTSKLDKYNKALNDLCIKVNNWDLFPSPSEDFTNDLNKGE 480
Qy 481 ITSDDTNIIEAENISLDDIQOYLYTFNFDNEPENISLENSSDIIGOLELMPNIERPNG 540
Db 481 ITSDDTNIIEAENISLDDIQOYLYTFNFDNEPENISLENSSDIIGOLELMPNIERPNG 540
Qy 541 KXYELDKYTMFHYLRAQEPFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 600
Db 541 KXYELDKYTMFHYLRAQEPFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNTGNMLYKDDFVGCALIFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNTGNMLYKDDFVGCALIFSG 660
Qy 661 AVILLEPEIAIPVLGTFFALVSIAKVLTVQTDNALSKRNEKWDDEVYKIIVTNWLAK 720
Db 661 AVILLEPEIAIPVLGTFFALVSIAKVLTVQTDNALSKRNEKWDDEVYKIIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCSSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Db 781 MININKFLNOCSSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Qy 841 VNTLSTDIIPFQLSKYVDNQRLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Db 841 VNTLSTDIIPFQLSKYVDNQRLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Qy 901 GSKVNFDPIDKNQIQLFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKNQIQLFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTIINCMMNSGKVSILNYGEIITWLTQDTEIKQVVFVKYSQMINISDYINRWIFVTIT 1020
Db 961 EYTIINCMMNSGKVSILNYGEIITWLTQDTEIKQVVFVKYSQMINISDYINRWIFVTIT 1020
Qy 1021 NNRLNNSKIYINGRLIDQKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRLNNSKIYINGRLIDQKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKOLYDNQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Db 1081 EKEIKOLYDNQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Qy 1141 GSVMTNIIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200

```

```

Db 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVVKKEYRLATNASQA 1200
Qy 1201 GVEKILSALETPDVGNLSQVVMKSKNDQGITNCKKNLQDNGNDIGFIFGHOFNNIAK 1260
Db 1201 GVEKILSALETPDVGNLSQVVMKSKNDQGITNCKKNLQDNGNDIGFIFGHOFNNIAK 1260
Qy 1261 LVASWYNRQIERSRSLTGCSEWFIPIVDDGNGERPL 1296
Db 1261 LVASWYNRQIERSRSLTGCSEWFIPIVDDGNGERPL 1296

RESULT 4
US-10-452-024-159
; Sequence 159, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Trans epithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 159
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-159

Query Match 99.2%; Score 6764; DB 15; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1294; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVNKQPNYKDPVNGVDIAIKIPNAGOMQPVKAFKIHKNKIWIPIERDFTNPEEGDLN 60
Db 1 MPFVNKQPNYKDPVNGVDIAIKIPNAGOMQPVKAFKIHKNKIWIPIERDFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYDSTYLTSTNEKDNVILKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPEAKQVPVSYDSTYLTSTNEKDNVILKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPGSYRSELNLVIIGPSADIIQFECKSGFHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPGSYRSELNLVIIGPSADIIQFECKSGFHEVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIIYAGHRLYGLAINPN 240
Db 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIIYAGHRLYGLAINPN 240
Qy 241 RVFKVNTNAYEMSGLEVSFELRTFGHDAKFIDSLQENEFRLYYNKKFKDIASTLNKA 300
Db 241 RVFKVNTNAYEMSGLEVSFELRTFGHDAKFIDSLQENEFRLYYNKKFKDIASTLNKA 300
Qy 301 KSIYGTASLOYMKNVKPKYLLSDETSKGFSVDKLFKDKLYKMLTEIYTEDNFVKPKV 360
Db 301 KSIYGTASLOYMKNVKPKYLLSDETSKGFSVDKLFKDKLYKMLTEIYTEDNFVKPKV 360
Qy 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRTNLAAFNQNGQNTENNMFVKLNKFT 420
Db 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRTNLAAFNQNGQNTENNMFVKLNKFT 420
Qy 421 GLFEPYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFPSPSDNFTNDLNKGE 480
Db 421 GLFEPYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFPSPSDNFTNDLNKGE 480
Qy 481 ITSDTNEAABENSLDIIQQYILTFNFDNPNENISLSSDIIQGLELMPNIERFPNG 540
Db 481 ITSDTNEAABENSLDIIQQYILTFNFDNPNENISLSSDIIQGLELMPNIERFPNG 540

```

```

Qy 541 KKVELDKYTMFHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFPSSDVYKKVKNKATEA 600
Db 541 KKVELDKYTMFHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFPSSDVYKKVKNKATEA 600
Qy 601 AMFLGWVEQLAVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMILYKDDFVGALIFSG 660
Db 601 AMFLGWVEQLAVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMILYKDDFVGALIFSG 660
Qy 661 AVILLBEPIEAIAPVLGTFFALVSIVANKVLTVQIDNALSKNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLBEPIEAIAPVLGTFFALVSIVANKVLTVQIDNALSKNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKKMEALENQAEATKAIINYQVQYTEBEKNININIDDLASKLNEINKA 780
Db 721 VNTQIDILIRKKMEALENQAEATKAIINYQVQYTEBEKNININIDDLASKLNEINKA 780
Qy 781 MININKELNOCSSVYLNMSIPIGVKLEDPDASLKDALLKYIYDNRGTLIGQVDRUKDK 840
Db 781 MININKELNOCSSVYLNMSIPIGVKLEDPDASLKDALLKYIYDNRGTLIGQVDRUKDK 840
Qy 841 VVNTLSTDIIPQLSKYVDNORLLSTFTEYIKNIINTSILMLRYESNHLIDLRYASKINI 900
Db 841 VVNTLSTDIIPQLSKYVDNORLLSTFTEYIKNIINTSILMLRYESNHLIDLRYASKINI 900
Qy 901 GSKVNFDPIDKNOIQLPNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKNOIQLPNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTIINCWENNSGWKYSNLNGEIIWTLODTEIKQORVVKYVSOMINISDYNRWIFVTIT 1020
Db 961 EYTIINCWENNSGWKYSNLNGEIIWTLODTEIKQORVVKYVSOMINISDYNRWIFVTIT 1020
Qy 1021 NNRLNNSKIYINGRLIDQKPIISNLGNHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRLNNSKIYINGRLIDQKPIISNLGNHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKDYDQNSGILKDFWGDYLOYDKPYMMLNLDPNKYVDVNNVGVIRGYMYLKGPR 1140
Db 1081 EKEIKDYDQNSGILKDFWGDYLOYDKPYMMLNLDPNKYVDVNNVGVIRGYMYLKGPR 1140
Qy 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVVKKEYRLATNASQA 1200
Db 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVVKKEYRLATNASQA 1200
Qy 1201 GVEKILSALETPDVGNLSQVVMKSKNDQGITNCKKNLQDNGNDIGFIFGHOFNNIAK 1260
Db 1201 GVEKILSALETPDVGNLSQVVMKSKNDQGITNCKKNLQDNGNDIGFIFGHOFNNIAK 1260
Qy 1261 LVASWYNRQIERSRSLTGCSEWFIPIVDDGNGERPL 1296
Db 1261 LVASWYNRQIERSRSLTGCSEWFIPIVDDGNGERPL 1296

RESULT 5
US-10-821-669-1
; Sequence 1, Application US/10821669
; Publication No. US20040265935A1
; GENERAL INFORMATION:
; APPLICANT: Atassi, M. Zouhair
; TITLE OF INVENTION: Botulinum Toxin A Peptides and Methods
; TITLE OF INVENTION: of Predicting and Reducing Immunoresistance to Botulinum
; FILE REFERENCE: 66872-039
; CURRENT APPLICATION NUMBER: US/10/821,669
; CURRENT FILING DATE: 2004-04-09
; PRIOR APPLICATION NUMBER: US 60/462,754
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum

```

US-10-821-669-1

Query Match 99.2%; Score 6764; DB 17; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1294; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVKQFNKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDTFTNPEEGDLN 60
Db 1 MPFVKQFNKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDTFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYDYSTYLDSTNEKDNLYKGYTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPEAKQVPVSYDYSTYLDSTNEKDNLYKGYTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
Db 241 RVFKVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIIVGTTASLOYMKQNVFKKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSIIVGTTASLOYMKQNVFKKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFDAVKAFKINIVKQNTIYDGNLRLNTLNAANFNGQNTINNMFVKLNFT 420
Db 361 LNRKTYLNFDAVKAFKINIVKQNTIYDGNLRLNTLNAANFNGQNTINNMFVKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKTSLDKGYKALNDLCIKVNNWDLFPSPEDNFTDNLNGEE 480
Db 421 GLFEFYKLLCVRGIIITSKTSLDKGYKALNDLCIKVNNWDLFPSPEDNFTDNLNGEE 480
Qy 481 ITSDNTIEAENISLDLIQOYIITFNFDNENISNIESDIIQOLELMPNTERFPNG 540
Db 481 ITSDNTIEAENISLDLIQOYIITFNFDNENISNIESDIIQOLELMPNTERFPNG 540
Qy 541 KKYELDKYTHPYLRAQEFEGKGRIALTNSVNEALLNPSRVYTFPSDYVKKYKATEA 600
Db 541 KKYELDKYTHPYLRAQEFEGKGRIALTNSVNEALLNPSRVYTFPSDYVKKYKATEA 600
Qy 601 AMFLGWEOQLVYDFTDETSEVSTTDKADIITIIPIYGPALNIGMLYKDDFVGCALIFSG 660
Db 601 AMFLGWEOQLVYDFTDETSEVSTTDKADIITIIPIYGPALNIGMLYKDDFVGCALIFSG 660
Qy 661 AVILLEFIPEITAIPLVGTGTFALVSIANKVLTQIDNALSKRNEKWDVYKIVTNWLAK 720
Db 661 AVILLEFIPEITAIPLVGTGTFALVSIANKVLTQIDNALSKRNEKWDVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKQWKALENQABATKAIINQVQYTEBEKNNINFIIDLSKLNESINKA 780
Db 721 VNTQIDILIRKQWKALENQABATKAIINQVQYTEBEKNNINFIIDLSKLNESINKA 780
Qy 781 MINTKFLNOCVSYLWNSMIPYGVKLEDFDASIKOALLKYIYDNRGTLIGQVDRUKDK 840
Db 781 MINTKFLNOCVSYLWNSMIPYGVKLEDFDASIKOALLKYIYDNRGTLIGQVDRUKDK 840
Qy 841 VNNTLSTDIPQLSKYVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900
Db 841 VNNTLSTDIPQLSKYVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900
Qy 901 GSKVNFDPIDKQIOLFNLSSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKQIOLFNLSSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTLINCENNSGWKSVLNYGEIITWLODQTEIKORVVVKYSQMINISDYINRWIFVTIT 1020
Db 961 EYTLINCENNSGWKSVLNYGEIITWLODQTEIKORVVVKYSQMINISDYINRWIFVTIT 1020

Qy 1021 NNRLLNSKIYINGRLIDQKPISNLGNHASNIMFKLDGCRDTHRYIWIYFNLFDEKLN 1080
Db 1021 NNRLLNSKIYINGRLIDQKPISNLGNHASNIMFKLDGCRDTHRYIWIYFNLFDEKLN 1080
Qy 1081 EKEIKOLYDNQNSGILKDFWGDYQYDKPYVYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Db 1081 EKEIKOLYDNQNSGILKDFWGDYQYDKPYVYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Qy 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA 1200
Db 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNCKRNLQDNNGNDIGFHFQFNIAK 1260
Db 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNCKRNLQDNNGNDIGFHFQFNIAK 1260
Qy 1261 LVASNWTNROIERSRSLTGLCSWEFIPVDDGWERPL 1296
Db 1261 LVASNWTNROIERSRSLTGLCSWEFIPVDDGWERPL 1296

RESULT 6

US-09-726-949A-1

; Sequence 1, Application US/09726949A
; Patent No. US20020137886A1
; GENERAL INFORMATION:
; APPLICANT: Allergan, Inc.
; APPLICANT: Lin, Wei-Jen
; APPLICANT: Aoki, Kei Roger
; APPLICANT: Steward, Lance E.
; TITLE OF INVENTION: NEUROTOXINS WITH ENHANCED TARGET
; TITLE OF INVENTION: SPECIFICITY
; FILE REFERENCE: 36121-20002.00
; CURRENT APPLICATION NUMBER: US/09/726,949A
; CURRENT FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1295
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-726-949A-1

Query Match 99.1%; Score 6759; DB 9; Length 1295;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PFVVKQFNKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDTFTNPEEGDLN 61
Db 1 PFVVKQFNKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDTFTNPEEGDLN 60
Qy 62 PPEAKQVPVSYDYSTYLDSTNEKDNLYKGYTKLPERIYSTDLGRMLLTSIVRGIPFWGG 121
Db 61 PPEAKQVPVSYDYSTYLDSTNEKDNLYKGYTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 122 TIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 181
Db 121 TIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 180
Qy 182 STQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIAPNP 241
Db 181 STQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIAPNP 240
Qy 242 VFQVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 301
Db 241 VFQVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
Qy 302 SIIVGTTASLOYMKQNVFKKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 361
Db 301 SIIVGTTASLOYMKQNVFKKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 360
Qy 362 NRTKTYLNFDAVKAFKINIVKQNTIYDGNLRLNTLNAANFNGQNTINNMFVKLNFT 421

361 NRKTYLNFDAVKFKNIVPKVNYTIDYGFNLNTNLAANFNGQNTENNMFNFKLKNFTG 420
422 LFPEYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 481
421 LFPEYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 480
482 TSDTNIEAABENISLDLIQQYLYTFNFDNPNENISLENLSSDIIGOLELMPNTERFNGK 541
481 TSDTNIEAABENISLDLIQQYLYTFNFDNPNENISLENLSSDIIGOLELMPNTERFNGK 540
542 KYELDKYTMFHYLRAQFEHGSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 601
541 KYELDKYTMFHYLRAQFEHGSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 600
602 MFLGWVQLVYDFDDETSEVSTTDKIADITIIPIYIGPALNIGNMVKDDPVGALIFSGA 661
601 MFLGWVQLVYDFDDETSEVSTTDKIADITIIPIYIGPALNIGNMVKDDPVGALIFSGA 660
662 VILLEFTEPEIAIPVLGTFALVSYANKVLTQTTIDNALSKNEKWDVYKIVTNWLAKV 721
661 VILLEFTEPEIAIPVLGTFALVSYANKVLTQTTIDNALSKNEKWDVYKIVTNWLAKV 720
722 NTQIDLIRKMKKEALENQAETKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 781
721 NTQIDLIRKMKKEALENQAETKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 780
782 ININKFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGLIGQVDRKDKV 841
781 ININKFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGLIGQVDRKDKV 840
842 NNTLSTDPFQLSKYVDNRQLSTFTYIKNIINTSILNLAYESNHLIDLSRYASKINIG 901
841 NNTLSTDPFQLSKYVDNRQLSTFTYIKNIINTSILNLAYESNHLIDLSRYASKINIG 900
902 SKVNFDPIDKQIQLFNLESSKIEVILKNAIVNYSMYENFSTFWIRIPKYFNSISLNE 961
901 SKVNFDPIDKQIQLFNLESSKIEVILKNAIVNYSMYENFSTFWIRIPKYFNSISLNE 960
962 YTIINCMMNSGKWSLNYGIIWTLODQEIQRVVPKYQSMINISDYINRWIFVTITN 1021
961 YTIINCMMNSGKWSLNYGIIWTLODQEIQRVVPKYQSMINISDYINRWIFVTITN 1020
1022 NRLANSKIYINGRLIDQKPI SNLGNHASNNIMFKLDCRDTYRVIWKYFNLFDKELNE 1081
1021 NRLANSKIYINGRLIDQKPI SNLGNHASNNIMFKLDCRDTYRVIWKYFNLFDKELNE 1080
1082 KEIKDLYDNQNSGILKDFWGDYLOYDKPYMLNLYDPNKVYDVNNVGIRGYMLKGRPG 1141
1081 KEIKDLYDNQNSGILKDFWGDYLOYDKPYMLNLYDPNKVYDVNNVGIRGYMLKGRPG 1140
1142 SVMTTNIYNSLSYRGTKFIKCVASGNKDNIVRNNDRVYINNVYKNEYRLATNASQAG 1201
1141 SVMTTNIYNSLSYRGTKFIKCVASGNKDNIVRNNDRVYINNVYKNEYRLATNASQAG 1200
1202 VEKILSALEIPDVGNLSQVVMKSKNDQGITNKCKMNLQDNNGNDIGFQHFQNNIAKL 1261
1201 VEKILSALEIPDVGNLSQVVMKSKNDQGITNKCKMNLQDNNGNDIGFQHFQNNIAKL 1260
1262 VASNNYNRQIERSRSTGLCSWEFIPVDDGWERPL 1296
1261 VASNNYNRQIERSRSTGLCSWEFIPVDDGWERPL 1295

RESULT 7
US-10-452-024-157
; Sequence 157, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transsepithelial Molecular Transport

FILE REFERENCE: 9855-96U1
CURRENT APPLICATION NUMBER: US/10/452,024
CURRENT FILING DATE: 2003-06-02
PRIOR APPLICATION NUMBER: 60/384,949
PRIOR FILING DATE: 2002-05-31
NUMBER OF SEQ ID NOS: 188
SOFTWARE: PatentIn version 3.2
SEQ ID NO 157
LENGTH: 1295
TYPE: PRT
ORGANISM: Clostridium botulinum
US-10-452-024-157

Query Match 99.1%; Score 6759; DB 15; Length 1295;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PFVNQPNYKDPNGVDIAYIKIPNAGOMQPVKAFKHNKIWVIPERDTNPEEGDLP 61
Db 1 PFVNQPNYKDPNGVDIAYIKIPNAGOMQPVKAFKHNKIWVIPERDTNPEEGDLP 60
Qy 62 PPEAKOVPSYDYSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLITSVIRGIPFWGGS 121
Db 61 PPEAKOVPSYDYSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLITSVIRGIPFWGGS 120
Qy 122 TIDTELKVIDTNCINVIQPDGYSRSEBELNLVIIGPSADIIQPECKSPGHVLMLTRNGYG 181
Db 121 TIDTELKVIDTNCINVIQPDGYSRSEBELNLVIIGPSADIIQPECKSPGHVLMLTRNGYG 180
Qy 182 STQVIRSPDFTGFPEESLVDNPLIGAGKFAATPAVTLAHELIVAGHRLYGIAINPNR 241
Db 181 STQVIRSPDFTGFPEESLVDNPLIGAGKFAATPAVTLAHELIVAGHRLYGIAINPNR 240
Qy 242 VFKNVNTAYEMSGLEVSFEELRTFGHDAKFIDSLQENEFRLYYNKKFDIASTLNKAK 301
Db 241 VFKNVNTAYEMSGLEVSFEELRTFGHDAKFIDSLQENEFRLYYNKKFDIASTLNKAK 300
Qy 302 SIVGTTASLOYMKNVFKKYLKLLSEDTSQKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKVL 361
Db 301 SIVGTTASLOYMKNVFKKYLKLLSEDTSQKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKVL 360
Qy 362 NRKTYLNFDAVKFKNIVPKVNYTIDYGFNLNTNLAANFNGQNTENNMFNFKLKNFTG 421
Db 361 NRKTYLNFDAVKFKNIVPKVNYTIDYGFNLNTNLAANFNGQNTENNMFNFKLKNFTG 420
Qy 422 LFPEYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 481
Db 421 LFPEYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 480
Qy 482 TSDTNIEAABENISLDLIQQYLYTFNFDNPNENISLENLSSDIIGOLELMPNTERFNGK 541
Db 481 TSDTNIEAABENISLDLIQQYLYTFNFDNPNENISLENLSSDIIGOLELMPNTERFNGK 540
Qy 542 KYELDKYTMFHYLRAQFEHGSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 601
Db 541 KYELDKYTMFHYLRAQFEHGSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 600
Qy 602 MFLGWVQLVYDFDDETSEVSTTDKIADITIIPIYIGPALNIGNMVKDDPVGALIFSGA 661
Db 601 MFLGWVQLVYDFDDETSEVSTTDKIADITIIPIYIGPALNIGNMVKDDPVGALIFSGA 660
Qy 662 VILLEFTEPEIAIPVLGTFALVSYANKVLTQTTIDNALSKNEKWDVYKIVTNWLAKV 721
Db 661 VILLEFTEPEIAIPVLGTFALVSYANKVLTQTTIDNALSKNEKWDVYKIVTNWLAKV 720
Qy 722 NTQIDLIRKMKKEALENQAETKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 781
Db 721 NTQIDLIRKMKKEALENQAETKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 780
Qy 782 ININKFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGLIGQVDRKDKV 841
Db 781 ININKFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGLIGQVDRKDKV 840

QY 842 NNTLSTDPFOLSKYVDNORLLSTFTEYIKNIINTSILNRYESNHLIDLSRYASKINIG 901
DB NNTLSTDPFOLSKYVDNORLLSTFTEYIKNIINTSILNRYESNHLIDLSRYASKINIG 900
QY 902 SKVNFDPIDKQIQIOLFNLESSKIEVLKNAIVNYSMYENFSTSWIRIPKYNFNSISLNE 961
DB SKVNFDPIDKQIQIOLFNLESSKIEVLKNAIVNYSMYENFSTSWIRIPKYNFNSISLNE 960
QY 962 YTIINCENNSGWKVSLSYGBIIWTLODTQBIKQVVPKYSQMINISDIYNRWIFVTITN 1021
DB YTIINCENNSGWKVSLSYGBIIWTLODTQBIKQVVPKYSQMINISDIYNRWIFVTITN 1020
QY 1022 NRLANSKIYINGRLIDQKPISNLGNHNASNNIMFKLDGCRDTHRYIWIKNFNLFDKELNE 1081
DB NRLANSKIYINGRLIDQKPISNLGNHNASNNIMFKLDGCRDTHRYIWIKNFNLFDKELNE 1080
QY 1082 KEIKDLDVQNSGILKDFWGDYLDYKPYPMNLXDPNKVDVNNVGIQYMYLKGPRG 1141
DB KEIKDLDVQNSGILKDFWGDYLDYKPYPMNLXDPNKVDVNNVGIQYMYLKGPRG 1140
QY 1142 SVMTTNIYLSLRYGTFKFIKKYASGNKONIVRRNDRVYINVVVKNEYRLATNASQAG 1201
DB SVMTTNIYLSLRYGTFKFIKKYASGNKONIVRRNDRVYINVVVKNEYRLATNASQAG 1200
QY 1202 VEKILSALEIPDVGNLSQVVMKSNDOGITNKCKMQLDNGNDIGFIGHQFNNTAKL 1261
DB VEKILSALEIPDVGNLSQVVMKSNDOGITNKCKMQLDNGNDIGFIGHQFNNTAKL 1260
QY 1262 VASWYNARQIERSRRTLGCSWEFIPVDDGWERPL 1296
DB VASWYNARQIERSRRTLGCSWEFIPVDDGWERPL 1295

RESULT 8

US-10-272-898-28
; Sequence 28, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-272-898-28
Query Match 99.0%; Score 6756; DB 14; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MPFVKQFNKYKDPVNGVDIAIYIKIPNAGQMPQKAFKIHKNKIWIPIERDFTFNEEGDLN 60
DB 1 MQFVKQFNKYKDPVNGVDIAIYIKIPNAGQMPQKAFKIHKNKIWIPIERDFTFNEEGDLN 60
QY 61 PPPEAKQPVSYSDYSTYLSTDNKKNYLKGVTKLFEIYSTDLGRMLLTSTVIRGIPFWGG 120
DB 61 PPPEAKQPVSYSDYSTYLSTDNKKNYLKGVTKLFEIYSTDLGRMLLTSTVIRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQFECKSGFGEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQFECKSGFGEVLNLTNGY 180
QY 181 GSTQYIRFSDPFTFGFEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP 240
DB 181 GSTQYIRFSDPFTFGFEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP 240
QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKPDKIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNKPDKIASTLNKA 300
QY 301 KSIIVGTTASLOYMKNVFEKYLLEDSTSGKFSVDKLFKDKLYKMLTETIYEDNFKVFPKV 360
DB 301 KSIIVGTTASLOYMKNVFEKYLLEDSTSGKFSVDKLFKDKLYKMLTETIYEDNFKVFPKV 360
QY 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGNLRLNTLAAFNFGONTINNNTFKLNFT 420
DB 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGNLRLNTLAAFNFGONTINNNTFKLNFT 420
QY 421 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480
DB 421 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480
QY 481 ITSDTNIEAAEENISLDLIQOYVLTFFNDEPENISLENSSDIIGOLEMNPINTERPNG 540
DB 481 ITSDTNIEAAEENISLDLIQOYVLTFFNDEPENISLENSSDIIGOLEMNPINTERPNG 540
QY 541 KYVELDKYTMFHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKYKVKATEA 600
DB 541 KYVELDKYTMFHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKYKVKATEA 600
QY 601 AMFLGWVEQLVYDFTDSTSVSTTDKIADITIIPIYIGPALNIGMMLYKDDFVGALIFSG 660
DB 601 AMFLGWVEQLVYDFTDSTSVSTTDKIADITIIPIYIGPALNIGMMLYKDDFVGALIFSG 660
QY 661 AVILLEFPIEPIAIPVLGTFFALVSIANKVLTQVOTIDNALSKRNEKWDVYKIYVNNLAK 720
DB 661 AVILLEFPIEPIAIPVLGTFFALVSIANKVLTQVOTIDNALSKRNEKWDVYKIYVNNLAK 720
QY 721 VNTQIDILIRKKMEALENOAEATKAIINYOQVQYTBEEKNNINFINIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKKMEALENOAEATKAIINYOQVQYTBEEKNNINFINIDDLSSKLNESINKA 780
QY 781 MININKFLNOCSSVYLMNSMIPYGVKRELEDFDASLKDALIKYIYDNRGTIGQVDRDKDK 840
DB 781 MININKFLNOCSSVYLMNSMIPYGVKRELEDFDASLKDALIKYIYDNRGTIGQVDRDKDK 840
QY 841 VNTLSTDIIPFOLSKYVDNORLLSTFTEYIKNIINTSILNRYESNHLIDLSRYASKINIG 900
DB 841 VNTLSTDIIPFOLSKYVDNORLLSTFTEYIKNIINTSILNRYESNHLIDLSRYASKINIG 900
QY 901 GSKVNFDPIDKQIQIOLFNLESSKIEVLKNAIVNYSMYENFSTSWIRIPKYNFNSISLNE 960
DB 901 GSKVNFDPIDKQIQIOLFNLESSKIEVLKNAIVNYSMYENFSTSWIRIPKYNFNSISLNE 960

QY 961 EYTIINCWENNSGKWSLNGEIIWTLDQTOEIKQKRVVFKYSQMINISDIYNRWIFVTIT 1020
DB 961 EYTIINCWENNSGKWSLNGEIIWTLDQTOEIKQKRVVFKYSQMINISDIYNRWIFVTIT 1020
QY 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
DB 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
QY 1081 EKEIKDLIDYQNSGIIKDFWGDYLDYDKPYMLNLYDPNKYVDVNNVGRGYWYLKGR 1140
DB 1081 EKEIKDLIDYQNSGIIKDFWGDYLDYDKPYMLNLYDPNKYVDVNNVGRGYWYLKGR 1140
QY 1141 GSVMTNINLSSLYRGTKFIKKYASGNKDNIRVNDNRVYINNVKNKEYRLATNASQA 1200
DB 1141 GSVMTNINLSSLYRGTKFIKKYASGNKDNIRVNDNRVYINNVKNKEYRLATNASQA 1200
QY 1201 GVEKILSALEIPDVGNIQSVVVMKSKNDQGITNKCKNVLQDNGNDIGFTGFQFNNAK 1260
DB 1201 GVEKILSALEIPDVGNIQSVVVMKSKNDQGITNKCKNVLQDNGNDIGFTGFQFNNAK 1260
QY 1261 LVASNNYRQIERSRSLTGCSEWFIPIVDDGNGERPL 1296
DB 1261 LVASNNYRQIERSRSLTGCSEWFIPIVDDGNGERPL 1296

RESULT 9

US-10-354-774-28
; Sequence 28, Application US/10354774
; Publication No. US20030215468A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Botulinum Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/354,774
; FILING DATE: 30-Jan-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane B.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-354-774-28

Query Match 99.0%; Score 6756; DB 14; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVNKQFNYPKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWVPERDTFTNPEEGDLN 60
DB 1 MQFVNKQFNYPKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWVPERDTFTNPEEGDLN 60
QY 61 PPPEAKQVPVSYDSTYLDNEKDYILKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
DB 61 PPPEAKQVPVSYDSTYLDNEKDYILKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIOPDGSYRSEELNLVIIGPSADIIIOPECKSFGEHVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIOPDGSYRSEELNLVIIGPSADIIIOPECKSFGEHVLNLTNGY 180
QY 181 GSTQYIRFSPDFTFGFEESLEVDNPLLAGKFPATDPAVTLAHELIIYAGHRLYGIAPN 240
DB 181 GSTQYIRFSPDFTFGFEESLEVDNPLLAGKFPATDPAVTLAHELIIYAGHRLYGIAPN 240
QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDKAFIDLSQENEFLLYYNFKDITASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDKAFIDLSQENEFLLYYNFKDITASTLNKA 300
QY 301 KSTVGTASLOQYMKNVFKKYLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
DB 301 KSTVGTASLOQYMKNVFKKYLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
QY 361 LNRKTYLNFDKAVFKINIYVKNYTIYDGFNLNTNLAANFNQONTNINMNFTKLKNFT 420
DB 361 LNRKTYLNFDKAVFKINIYVKNYTIYDGFNLNTNLAANFNQONTNINMNFTKLKNFT 420
QY 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNWDLFPSPEDNFTNDLNGEE 480
DB 421 GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNWDLFPSPEDNFTNDLNGEE 480
QY 481 ITSDTNIEAAEENISLDLIQOYYLTFNFDNEPENISIENLSSDIIGQELMPNIEFPNG 540
DB 481 ITSDTNIEAAEENISLDLIQOYYLTFNFDNEPENISIENLSSDIIGQELMPNIEFPNG 540
QY 541 KKYELDKYTMFHYLRAQEPFHGKSRITLTVSNVNEALLNPSRVVTFPSSDYVKKVKNKATEA 600
DB 541 KKYELDKYTMFHYLRAQEPFHGKSRITLTVSNVNEALLNPSRVVTFPSSDYVKKVKNKATEA 600
QY 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDDFVGLIFSG 660
DB 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDDFVGLIFSG 660
QY 661 AVILLEPIEIAIPVLGTALVSVIANKVLTQVOTIDNLSKRNEKMDDEVYKIVTNWLAK 720
DB 661 AVILLEPIEIAIPVLGTALVSVIANKVLTQVOTIDNLSKRNEKMDDEVYKIVTNWLAK 720
QY 721 VNTQIDILIRKMKALENOQAEATKAIINYQNYQTEEEKNNINFNIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKMKALENOQAEATKAIINYQNYQTEEEKNNINFNIDDLSSKLNESINKA 780
QY 781 MININKELNOCVSYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGVDRLKDK 840
DB 781 MININKELNOCVSYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTILIGVDRLKDK 840
QY 841 VVNTLSTDIDPFQLSKYVDNORLLSTFTTEYIKNIINTSILNRYESNHLIDSLRYASKINI 900
DB 841 VVNTLSTDIDPFQLSKYVDNORLLSTFTTEYIKNIINTSILNRYESNHLIDSLRYASKINI 900
QY 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVVNSMYENFSTFWIRIPKYFNSISLNN 960
DB 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVVNSMYENFSTFWIRIPKYFNSISLNN 960
QY 961 EYTIINCWENNSGKWSLNGEIIWTLDQTOEIKQKRVVFKYSQMINISDIYNRWIFVTIT 1020
DB 961 EYTIINCWENNSGKWSLNGEIIWTLDQTOEIKQKRVVFKYSQMINISDIYNRWIFVTIT 1020
QY 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
DB 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080

Qy	1201	GVEKILSALIEIPDVGNLSQVVVMKSKNDQGITNKCCKNMLDNNKNDIGFTGFHFQFNIAK	1260
Db	1201	GVEKILSALIEIPDVGNLSQVVVMKSKNDQGITNKCCKNMLDNNKNDIGFTGFHFQFNIAK	1260
Qy	1261	LVASNNWYNROIERSRSLTGCWSWEIFPVDDGNGERPL	1296
Db	1261	LVASNNWYNROIERSRSLTGCWSWEIFPVDDGNGERPL	1296

RESULT 11

```

US-10-452-024-158
; Sequence 156, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 156
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-158

```

Query Match	99.0%;	Score 6756;	DB 15;	Length 1296;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1293;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1	MPFVKNQFNYPKDPVNGVDIAIYKIPNAGQMPVKAFKIHNKIWIIPERDTEFNPEEGDLN	60	
Db	1	MQFVKNQFNYPKDPVNGVDIAIYKIPNVGQMPVKAFKIHNKIWIIPERDTEFNPEEGDLN	60	
Qy	61	PPPEAKQVPVSYDYSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVIRGIPFWGG	120	
Db	61	PPPEAKQVPVSYDYSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVIRGIPFWGG	120	
Qy	121	STIDTELKVIDTNCINVIQPDGSYRSBELNLVIIGPSADIIOFCKSGFGEVNLNTRNGY	180	
Db	121	STIDTELKVIDTNCINVIQPDGSYRSBELNLVIIGPSADIIOFCKSGFGEVNLNTRNGY	180	
Qy	181	GSTQYIRFSPDFTFGFRESLEVDNPNLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP	240	
Db	181	GSTQYIRFSPDFTFGFRESLEVDNPNLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP	240	
Qy	241	RVFKVNTNAYYEMSGLEVSPEELATFGCHDAKFIDSLOQENEFRIYYNKKFDIASTLNKA	300	
Db	241	RVFKVNTNAYYEMSGLEVSPEELATFGCHDAKFIDSLOQENEFRIYYNKKFDIASTLNKA	300	
Qy	301	KSIVGTTASLQYMKNVFKCKYLLSEDTSKGFSVDKLPDKLYKMLTEIYTDENFVKPFKV	360	
Db	301	KSIVGTTASLQYMKNVFKCKYLLSEDTSKGFSVDKLPDKLYKMLTEIYTDENFVKPFKV	360	
Qy	361	LNRTYLNFDKAVPKINIVPKVNTIYDGFNLRTNLNLAANFNGQNTIINNNTFKLNKFT	420	
Db	361	LNRTYLNFDKAVPKINIVPKVNTIYDGFNLRTNLNLAANFNGQNTIINNNTFKLNKFT	420	
Qy	421	GLFIFYKLLCVRGIIITSKTSLSLDKGYNKALNDLCIKVNNWDLFFSPSDFNTDLNKGEE	480	
Db	421	GLFIFYKLLCVRGIIITSKTSLSLDKGYNKALNDLCIKVNNWDLFFSPSDFNTDLNKGEE	480	
Qy	481	ITSDTNIIEAEEENTSLDLIOOYLYTFNFDNEPENISIENLSSDIIGOLELMPNIEFPNG	540	
Db	481	ITSDTNIIEAEEENTSLDLIOOYLYTFNFDNEPENISIENLSSDIIGOLELMPNIEFPNG	540	
Qy	541	KKYBELDKYTMPIYLRABEFHGKGRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600	

541	Db	541	KKYELDKYTMFWHYLRAQEPHGHKSRJALTJNSVNEALLNPSRVVTFPSSDYVVKVKNKATEA	600
601	Qy	601	AMFLGWQVLQYDFDTDETSEVSTDKIADITIIIPYIGPALNITGNMLYKDDFVGALIFSG	660
601	Db	601	AMFLGWQVLQYDFDTDETSEVSTDKIADITIIIPYIGPALNITGNMLYKDDFVGALIFSG	660
661	Qy	661	AVILLFPIEPIAIPVLGTFAVSYIAKULVTQOTIDNALSKRNEKWDEVYKYIVTNWLAK	720
661	Db	661	AVILLFPIEPIAIPVLGTFAVSYIAKULVTQOTIDNALSKRNEKWDEVYKYIVTNWLAK	720
721	Qy	721	VNTQIDILIRKMKKEALENQAEATKAIINYOQNOYTBEEKNNINFNIDDLSSKLNESINKA	780
721	Db	721	VNTQIDILIRKMKKEALENQAEATKAIINYOQNOYTBEEKNNINFNIDDLSSKLNESINKA	780
781	Qy	781	MININKFLNQCYSVYLMNSMIPYGVKRLKEDFASLKDALLKYIYDNRGTILIGQVDRUKOK	840
781	Db	781	MININKFLNQCYSVYLMNSMIPYGVKRLKEDFASLKDALLKYIYDNRGTILIGQVDRUKOK	840
841	Qy	841	VNNTLSTDIPFQLSKYVDNORLLSTFTEYIKNIINTSIILNRYESNHLIDLSRYASKINI	900
841	Db	841	VNNTLSTDIPFQLSKYVDNORLLSTFTEYIKNIINTSIILNRYESNHLIDLSRYASKINI	900
901	Qy	901	GSKVNFDPDKNOIQLFNLESSKIEVLKNAIIVYNSMYENFSTFWIRIBKYFNFSISLNN	960
901	Db	901	GSKVNFDPDKNOIQLFNLESSKIEVLKNAIIVYNSMYENFSTFWIRIBKYFNFSISLNN	960
961	Qy	961	EYTIINCMMNNSGWKYSVLYNGEIIWTLODQTEIKQRVFKYSOMINISDIYNENWIFVTIT	1020
961	Db	961	EYTIINCMMNNSGWKYSVLYNGEIIWTLODQTEIKQRVFKYSOMINISDIYNENWIFVTIT	1020
1021	Qy	1021	NNRLNNSKIYINGRLIDQKPISNLGNITHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELN	1080
1021	Db	1021	NNRLNNSKIYINGRLIDQKPISNLGNITHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELN	1080
1081	Qy	1081	EKEIKOLYDNQNSGILKOPFGWBYLOYDKPYMYMLNLYDPNKYVDVNVNGIRGYMYLKGPR	1140
1081	Db	1081	EKEIKOLYDNQNSGILKOPFGWBYLOYDKPYMYMLNLYDPNKYVDVNVNGIRGYMYLKGPR	1140
1141	Qy	1141	GSVYMTNIIYLNSSIRYKFTKPIIKKYASGNKDNIVRNNDRVYINVVYKNGEYRLATNASQA	1200
1141	Db	1141	GSVYMTNIIYLNSSIRYKFTKPIIKKYASGNKDNIVRNNDRVYINVVYKNGEYRLATNASQA	1200
1201	Qy	1201	GVEKILSALSAEIPDVGNLSQVVMVMSKNDQGITNKCKMNLQDNGNDIGFIGHQFNNI	1260
1201	Db	1201	GVEKILSALSAEIPDVGNLSQVVMVMSKNDQGITNKCKMNLQDNGNDIGFIGHQFNNI	1260
1261	Qy	1261	LVASNMYNROIERSRSTLGCWSBFIPVDDGWBGRPL	1296
1261	Db	1261	LVASNMYNROIERSRSTLGCWSBFIPVDDGWBGRPL	1296

RESULT 12

RES001.12
US-10-729-122-28
Sequence 28, Application US/10729122
Publication No. US20040115215A1
GENERAL INFORMATION:
APPLICANT: Williams, James A.
Thalley, Bruce S.
TITLE OF INVENTION: Bivalent Vaccine For Clostridium Botulinum Neurotoxin
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-729-039-28

Query Match 99.08; Score 6756; DB 16; Length 1296;
Best Local Similarity 99.84; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVVKQPNYKDPVNGVDIAYIKIPNAGQMPKAFKIHKNKIWIPIPERDFTNPEEGDLN 60
DB 1 MQFVVKQPNYKDPVNGVDIAYIKIPNAGQMPKAFKIHKNKIWIPIPERDFTNPEEGDLN 60
QY 61 PPPEAKQPVSYDSTYLSSTDNKDKYLGKVTCLFERIYSTDLGRMLLTSIVRGIPFWGG 120
DB 61 PPPEAKQPVSYDSTYLSSTDNKDKYLGKVTCLFERIYSTDLGRMLLTSIVRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQDGSYRSELNLVIIGPSADIIQPECKSFGEHVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQDGSYRSELNLVIIGPSADIIQPECKSFGEHVLNLTNGY 180
QY 181 GSTQYIRFSPDFTFGFESLEVDNPLLAGKFAFDPAVTLAHLIYAGHRLYGLIAPNP 240
DB 181 GSTQYIRFSPDFTFGFESLEVDNPLLAGKFAFDPAVTLAHLIYAGHRLYGLIAPNP 240
QY 241 RVFKVNTNAYEMSGLEVSFEELFTFGHDHAKFTDSLOENEFRLYYNKKFDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELFTFGHDHAKFTDSLOENEFRLYYNKKFDIASTLNKA 300
QY 301 KSIIVGTTASQYKMNPKFKYLLSEDTSGKFSVDKLFKLYKMLTIYTEDNFKVFKV 360
DB 301 KSIIVGTTASQYKMNPKFKYLLSEDTSGKFSVDKLFKLYKMLTIYTEDNFKVFKV 360
QY 361 LNRKTYLNFDFKAVKINIPKVNVTIYDGFNLRLNTLAAFPNGQNTENNMFVKLNFT 420
DB 361 LNRKTYLNFDFKAVKINIPKVNVTIYDGFNLRLNTLAAFPNGQNTENNMFVKLNFT 420
QY 421 GLFEPYKLLCVRGITTSKTSKLDKGYKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
DB 421 GLFEPYKLLCVRGITTSKTSKLDKGYKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
QY 481 ITSDNTNIEAENISLDLIQYVLTFFNFPENISLENLSSDIIGOLELMPNTERPPNG 540
DB 481 ITSDNTNIEAENISLDLIQYVLTFFNFPENISLENLSSDIIGOLELMPNTERPPNG 540
QY 541 KKYELDKYTMFHYLRAQFEHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
DB 541 KKYELDKYTMFHYLRAQFEHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
QY 601 AMFLGWVQLVYDFDTESEVSTTDKIADITIIPIYIGPALNIGNMLYKDFVGLIFSG 660
DB 601 AMFLGWVQLVYDFDTESEVSTTDKIADITIIPIYIGPALNIGNMLYKDFVGLIFSG 660
QY 661 AVILLEFTPEIAPVLGTFALVSYTANKVLTQVIDNALSKRNEKWDVYKVIYVNNMLAK 720
DB 661 AVILLEFTPEIAPVLGTFALVSYTANKVLTQVIDNALSKRNEKWDVYKVIYVNNMLAK 720
QY 721 VNTQIDLIRKKMKEALENOAETKAIINQYQNTYEEKNNINFNIDLSKLNESINKA 780
DB 721 VNTQIDLIRKKMKEALENOAETKAIINQYQNTYEEKNNINFNIDLSKLNESINKA 780
QY 781 MININKFLNOCQSVYLMNSMIPYGVKRLDFDASLKDALLYKVIYDNRGTLIGQVDELKDK 840

781 MININKFLNOCQSVYLMNSMIPYGVKRLDFDASLKDALLYKVIYDNRGTLIGQVDELKDK 840
QY 841 VNNLTSTDIPFQLSKYVDNQRLLSITFETIYKNIINTSIILNRYESNHLIDLRYASKINI 900
DB 841 VNNLTSTDIPFQLSKYVDNQRLLSITFETIYKNIINTSIILNRYESNHLIDLRYASKINI 900
QY 901 GSKVNFDPIDKQIQOLFNLSESSKIEVILKNAIVNYSMYENFSTSFWRIPKYPFNSISLNN 960
DB 901 GSKVNFDPIDKQIQOLFNLSESSKIEVILKNAIVNYSMYENFSTSFWRIPKYPFNSISLNN 960
QY 961 EYTIINCMMNSGKVSLSNYGELIITWLTQTOEIKQVWFKYSQMINISYINRWIFVTIT 1020
DB 961 EYTIINCMMNSGKVSLSNYGELIITWLTQTOEIKQVWFKYSQMINISYINRWIFVTIT 1020
QY 1021 NNRLNNSKIYINGRLIDQKPISNLGNHASNIMFKLDCGRDTHRYIWIYKYNLFDKELN 1080
DB 1021 NNRLNNSKIYINGRLIDQKPISNLGNHASNIMFKLDCGRDTHRYIWIYKYNLFDKELN 1080
QY 1081 EKEIKDLYDNQNSGILKDFWGDYLOQDKPYMYMLNLYDNKYVDVNNVGIRGYMYLKGP 1140
DB 1081 EKEIKDLYDNQNSGILKDFWGDYLOQDKPYMYMLNLYDNKYVDVNNVGIRGYMYLKGP 1140
QY 1141 GSVMTTNIYLNSSLYRGTFIIFKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
DB 1141 GSVMTTNIYLNSSLYRGTFIIFKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
QY 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKCQKQNDQNNNDIGIFGFHGFNNIAK 1260
DB 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKCQKQNDQNNNDIGIFGFHGFNNIAK 1260
QY 1261 LVASNNYNRQIERSRSTLGCSEWFEFIPVDGWSERPL 1296
DB 1261 LVASNNYNRQIERSRSTLGCSEWFEFIPVDGWSERPL 1296

RESULT 14
US-10-729-527-28
; Sequence 28, Application US/10729527
; Publication No. US20040219637A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thallev, Bruce S.
; TITLE OF INVENTION: Botulinum Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESS: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10729,527
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:

;	;	LENGTH: 1296 amino acids			
;	;	TOPOLOGY: linear			
;	;	MOLECULE TYPE: protein			
;	;	SEQUENCE DESCRIPTION: SEQ ID NO: 28:			
US-10-729-527-28					
Query Match 99.0%; Score 6756; DB 17; Length 1296;					
Best Local Similarity 99.8%; Pred. No. 0;					
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;					
QY	1	MPFVKQFNKQDPVNGVDIAIKIPNAGQMPKAFKIHKNKIWIIPEDRTFTNPEEGDLN	60		
DB	1	MQFVKQFNKQDPVNGVDIAIKIPNAGQMPKAFKIHKNKIWIIPEDRTFTNPEEGDLN	60		
QY	61	PPPEAKQPVSYDYSTYLSSTDEKDKYLKGVTKLFEIYSTDGLRMLLTISVIRGIPFWGG	120		
DB	61	PPPEAKQPVSYDYSTYLSSTDEKDKYLKGVTKLFEIYSTDGLRMLLTISVIRGIPFWGG	120		
QY	121	STIDTELKVIDTNCINVIQDGSYRSEELNLVIIGPSADIIQPECKSFGEHVLNLTNGY	180		
DB	121	STIDTELKVIDTNCINVIQDGSYRSEELNLVIIGPSADIIQPECKSFGEHVLNLTNGY	180		
QY	181	GSTQYIRFSPDFTGPFESLEVDNPLLGAGKFAATDPAVTLAHLIYAGHRLYGIAPNP	240		
DB	181	GSTQYIRFSPDFTGPFESLEVDNPLLGAGKFAATDPAVTLAHLIYAGHRLYGIAPNP	240		
QY	241	RVKQVNTNAYYEMSGLEVPSELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA	300		
DB	241	RVKQVNTNAYYEMSGLEVPSELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA	300		
QY	301	KSIIVGTTASLOYQMNVPKFKYLLSEDTSGKFSVDKLPKDLKYLKMLTEIYTEDNPFVKPV	360		
DB	301	KSIIVGTTASLOYQMNVPKFKYLLSEDTSGKFSVDKLPKDLKYLKMLTEIYTEDNPFVKPV	360		
QY	361	LNKRTYLNFKAVKINIVPKVNTIYDGNFLRNTNLAANFNGQNTINNPNFTKLKNT	420		
DB	361	LNKRTYLNFKAVKINIVPKVNTIYDGNFLRNTNLAANFNGQNTINNPNFTKLKNT	420		
QY	421	GLPEFYKLLCVRGIIITSKLSLDKGVKALNDLCIKVNNWDLFPSSEDFNFTDLNKGEE	480		
DB	421	GLPEFYKLLCVRGIIITSKLSLDKGVKALNDLCIKVNNWDLFPSSEDFNFTDLNKGEE	480		
QY	481	ITSDNTIEAABENISLDLIQYLTFFNPNENPISNIENSSDIIGOLELMPNIERFPNG	540		
DB	481	ITSDNTIEAABENISLDLIQYLTFFNPNENPISNIENSSDIIGOLELMPNIERFPNG	540		
QY	541	KYIELDKYTWPHYLRAQEFPHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600		
DB	541	KYIELDKYTWPHYLRAQEFPHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600		
QY	601	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGCALIFSG	660		
DB	601	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGCALIFSG	660		
QY	661	AVILLEFPIETAIPIVLTGTFALVSIAKVLTVQTDNALSKRNEKWDVYKVIYVNWLAKE	720		
DB	661	AVILLEFPIETAIPIVLTGTFALVSIAKVLTVQTDNALSKRNEKWDVYKVIYVNWLAKE	720		
QY	721	VNTQIDILIRKQKALENQAETAKAIINYQVQNTTEBEKNININIDDLSSKLNESINKA	780		
DB	721	VNTQIDILIRKQKALENQAETAKAIINYQVQNTTEBEKNININIDDLSSKLNESINKA	780		
QY	781	MININKFLNOCVSYLNSMIPYGVKLEDPDASLKDALIKYIYDNRGTGLIGQVDRUKDK	840		
DB	781	MININKFLNOCVSYLNSMIPYGVKLEDPDASLKDALIKYIYDNRGTGLIGQVDRUKDK	840		
QY	841	VNNTLSTDIIPQLSKYVDNORLLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI	900		
DB	841	VNNTLSTDIIPQLSKYVDNORLLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI	900		
QY	901	GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVTNSMYENFSTFWIRIPKYNFNSISLNN	960		

Db	901	GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVTNSMYENFSTFWIRIPKYNFNSISLNN	960		
QY	961	EYTIINCMMNSGWKVSILNYGEIITWLTQDTQEIQRVVFKYSQMINISDYINRWIFVTIT	1020		
Db	961	EYTIINCMMNSGWKVSILNYGEIITWLTQDTQEIQRVVFKYSQMINISDYINRWIFVTIT	1020		
QY	1021	NNRLNNSKIYINGRLIDOKPISNLGNIIHASNNIMPKLDGCRDTHRYIWKYPNLPDKELN	1080		
Db	1021	NNRLNNSKIYINGRLIDOKPISNLGNIIHASNNIMPKLDGCRDTHRYIWKYPNLPDKELN	1080		
QY	1081	EKEIKDLYDNQNSGILKDFWGDYLOQDPKPYMNLNLYDPNKYVDVNNVGIRGYMYLKGPR	1140		
Db	1081	EKEIKDLYDNQNSGILKDFWGDYLOQDPKPYMNLNLYDPNKYVDVNNVGIRGYMYLKGPR	1140		
QY	1141	GSVMTTNIYNSLVRGTFKFIKKYASGNKONIVRNDRVYINVVVKNYKRYLATNASQA	1200		
Db	1141	GSVMTTNIYNSLVRGTFKFIKKYASGNKONIVRNDRVYINVVVKNYKRYLATNASQA	1200		
QY	1201	GVEKILSALEIPDVNLSQVVMKSKNDQGITNCKMNLQDNNGNDIGFIGFHQFNNTAK	1260		
Db	1201	GVEKILSALEIPDVNLSQVVMKSKNDQGITNCKMNLQDNNGNDIGFIGFHQFNNTAK	1260		
QY	1261	LVASNNYNRQIERSRTILGCSWEFIPVDVDDGWERPL	1296		
Db	1261	LVASNNYNRQIERSRTILGCSWEFIPVDVDDGWERPL	1296		
RESULT 15					
US-10-727-898-28					
; Sequence 28, Application US/10727898					
; Publication No. US20040235118A1					
; GENERAL INFORMATION:					
; APPLICANT: Williams, James A.					
; Thallev, Bruce S.					
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium					
; Botulinum Neurotoxin					
; NUMBER OF SEQUENCES: 82					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: Medlen & Carroll					
; STREET: 220 Montgomery Street, Suite 2200					
; CITY: San Francisco					
; STATE: California					
; COUNTRY: United States of America					
; ZIP: 94104					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Floppy disk					
; OPERATING SYSTEM: PC compatible					
; SOFTWARE: Patentin Release #1.0, Version #1.30					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/10/727,898					
; FILING DATE: 04-Dec-2003					
; CLASSIFICATION: 424					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US/08/704,159					
; FILING DATE: <Unknown>					
; ATTORNEY/AGENT INFORMATION:					
; NAME: Ingolia, Diane E.					
; REGISTRATION NUMBER: 40,027					
; REFERENCE/DOCKET NUMBER: OPHD-02304					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: (415) 705-8410					
; TELEFAX: (415) 397-8338					
; INFORMATION FOR SEQ ID NO: 28:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 1296 amino acids					
; TYPE: amino acid					
; TOPOLOGY: linear					
; MOLECULE TYPE: protein					
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:					
US-10-727-898-28					
Query Match 99.0%; Score 6756; DB 17; Length 1296;					

Query Match

99.0%; Score 6756; DB 17; Length 1296;

Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;			
Qy	1	MPFVNFQFNKDPVNGVDIAYIKIPNAGQMPVKAFAKIHKNKIWIIPERDFTFTPEEGDLN	60
Db	1	MQFVNFQFNKDPVNGVDIAYIKIPNAGQMPVKAFAKIHKNKIWIIPERDFTFTPEEGDLN	60
Qy	61	PPPEAKQVPVSYDSTYLSSTNEDKDYLGVTYKLFERIYSTDLGRMLLTISVIRGIPFWGG	120
Db	61	PPPEAKQVPVSYDSTYLSSTNEDKDYLGVTYKLFERIYSTDLGRMLLTISVIRGIPFWGG	120
Qy	121	STIDTELKVIDTNCINVIQPDGSYRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY	180
Db	121	STIDTELKVIDTNCINVIQPDGSYRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY	180
Qy	181	GSTQYIRFSPDFTFGFESLEVDNPLLGAKFATDPAVTLAHELIYAGHRLYGIAPNP	240
Db	181	GSTQYIRFSPDFTFGFESLEVDNPLLGAKFATDPAVTLAHELIYAGHRLYGIAPNP	240
Qy	241	RVFKVNTNAYYEMSGLEVSFEELRTFGHDAKFIDSLEQENEFRLYYNKKFDIASTLKA	300
Db	241	RVFKVNTNAYYEMSGLEVSFEELRTFGHDAKFIDSLEQENEFRLYYNKKFDIASTLKA	300
Qy	301	KSIVGTTASLOYMKVFKPKYLLSEDTSKPSVDKLPDKLYKMLTEIYTEDNFVKPFKV	360
Db	301	KSIVGTTASLOYMKVFKPKYLLSEDTSKPSVDKLPDKLYKMLTEIYTEDNFVKPFKV	360
Qy	361	LNRKTYLNFDAVKFKNIVPKVNTIYDGNLRLNTLAAFPNGONTNINNNFTKLKNFT	420
Db	361	LNRKTYLNFDAVKFKNIVPKVNTIYDGNLRLNTLAAFPNGONTNINNNFTKLKNFT	420
Qy	421	GLFEFYKLLCVRGIIITSKTSKLDGYNKALNDLCIKVNNWDLFPSSEDNFTDNLKGE	480
Db	421	GLFEFYKLLCVRGIIITSKTSKLDGYNKALNDLCIKVNNWDLFPSSEDNFTDNLKGE	480
Qy	481	ITSDTNEAABENTSLDIIQOYLYITFNPDPENISIENTSSDIIGOLELMPNTERPFG	540
Db	481	ITSDTNEAABENTSLDIIQOYLYITFNPDPENISIENTSSDIIGOLELMPNTERPFG	540
Qy	541	KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600
Db	541	KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600
Qy	601	AMFLGWVQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKODFVGALIFSG	660
Db	601	AMFLGWVQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKODFVGALIFSG	660
Qy	661	AVILLEFIPEIAIPVLGTFALVSYIANKVLTVOIDNALSKRNEKWDVYKYIIVTNWLAK	720
Db	661	AVILLEFIPEIAIPVLGTFALVSYIANKVLTVOIDNALSKRNEKWDVYKYIIVTNWLAK	720
Qy	721	VNTQIDLLRKWKKEALENQAEATKAIINYQNYTEEBEKNINFINIDDLSSKLNESINKA	780
Db	721	VNTQIDLLRKWKKEALENQAEATKAIINYQNYTEEBEKNINFINIDDLSSKLNESINKA	780
Qy	781	MININKFLNQCYSVLYMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGQVDRLKDK	840
Db	781	MININKFLNQCYSVLYMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGQVDRLKDK	840
Qy	841	VNNTLSTDIIPFOLSKYVDNQBLSTFTTEYIKNIINTSILNRYESNHLIDLSRYASKINI	900
Db	841	VNNTLSTDIIPFOLSKYVDNQBLSTFTTEYIKNIINTSILNRYESNHLIDLSRYASKINI	900
Qy	901	GSKVNFDPIDKNOIQLFNLESKEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN	960
Db	901	GSKVNFDPIDKNOIQLFNLESKEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN	960
Qy	961	EYTIINCINNENSGWKVSLNYGEIITLQDTQEIQRVVFYKYSQMINISDYINRWIFVTIT	1020
Db	961	EYTIINCINNENSGWKVSLNYGEIITLQDTQEIQRVVFYKYSQMINISDYINRWIFVTIT	1020
Qy	1021	NNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN	1080

Search completed: January 31, 2005, 14:12:58
Job time : 181 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:03 ; Search time 52 Seconds
(without alignments)
2409.120 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFVNQFNKYDVPNGVDIA.....EFIPVDGWERPLHHHHH 1302

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6768	99.2	1296	1 BTCLAB	bontoxilysin (EC 3
2	6158	90.3	1296	2 I40645	botulinum neurotox
3	2504.5	36.7	1268	2 S33411	botulinum neurotox
4	2339	34.3	1252	2 S21117	botulinum neurotox
5	2324.5	34.1	1251	2 JH0256	botulinum neurotox
6	2296.5	33.7	1274	2 I40813	neurotoxin type F
7	2285	33.5	1297	2 S39791	neurotoxin - Clost
8	2271	33.3	1291	1 A48940	bontoxilysin (EC 3
9	2267	33.2	1291	2 I40631	non-proteolytic bo
10	1892.5	27.7	1315	1 BTCLTN	tentoxilysin (EC 3
11	1890.5	27.7	1276	2 S11455	botulinum neurotox
12	1851	27.1	1285	2 S70582	botulinum neurotox
13	1771	26.0	1291	2 A49777	botulinum neurotox
14	1771	26.0	1291	2 S46431	botulinum neurotox
15	847	12.4	366	2 S48110	neurotoxin type F
16	805.5	11.8	367	2 S48106	neurotoxin type E
17	804.5	11.8	367	2 S48109	neurotoxin type F
18	616	9.0	552	2 A49928	neurotoxin Cl, BoN
19	614.5	9.0	1162	2 A47708	progenitor toxin n
20	614.5	9.0	1162	2 I40817	botulinum toxin no
21	583	8.5	1193	2 S68218	botulinum neurotox
22	567.5	8.3	1196	2 J01467	toxin, nontoxic co
23	566.5	8.3	1196	2 S64430	botulinum neurotox
24	553.5	8.1	1193	2 J04901	nontoxic-nonhemagg
25	534.5	7.8	1165	2 I40644	botulinum neurotox
26	318.5	4.7	2401	2 T28676	roptory protein -
27	310	4.5	960	2 S72284	DNA-directed RNA p
28	302.5	4.4	2269	2 T28677	roptory protein -
29	297.5	4.4	1127	2 T28317	ORF MSV156 hypothe

30	280	4.1	3724	2	T18427	hypothetical prote
31	280	4.1	4550	2	T18440	hypothetical prote
32	276.5	4.1	2485	1	H71621	serine/threonine-s
33	276.5	4.1	4688	2	F82885	hypothetical prote
34	272.5	4.0	1830	2	E82909	conserved hypotet
35	268	3.9	3394	2	T18501	hypothetical prote
36	267	3.9	2819	2	A90551	conserved hypotet
37	265	3.9	2380	2	E71604	hypothetical prote
38	264.5	3.9	4981	2	T18489	hypothetical prote
39	264	3.9	1183	2	F90559	conserved hypotet
40	261.5	3.8	2539	2	B71619	hypothetical prote
41	259.5	3.8	1121	2	F71613	hypothetical prote
42	258.5	3.8	1447	2	F82909	hypothetical prote
43	257.5	3.8	1346	2	G71613	hypothetical prote
44	257	3.8	5005	2	F82884	hypothetical prote
45	256.5	3.8	2136	2	A05037	hypothetical prote

ALIGNMENTS

RESULT 1

BTCLAB

bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum

N;Alternate names: botulinum neurotoxin type A

C;Species: Clostridium botulinum

C;Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text change 09-Jul-2004

C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000

R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.

J. Biol. Chem. 265, 9153-9158, 1990

A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other

A;Reference number: A35294; MUID:90264400; PMID:2160960

A;Accession: A35294

A;Molecule type: DNA

A;Residues: 1-1296 <BIN>

A;Cross-references: UNIPROT:P10845; GB:M30196; NID:G144864; PIDN:AAA23262.1; PID:G144865

A;Experimental source: strain 62A, subtype A

R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.

Eur. J. Biochem. 189, 73-81, 1990

A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin

A;Reference number: S09492; MUID:90235864; PMID:2185020

A;Accession: S09492

A;Molecule type: DNA

A;Residues: 1, 'Q', 3-26, 'V', 28-1296 <THO>

A;Cross-references: EMBL:X52066; NID:G40381; PIDN:CAA36289.1; PID:G40382

A;Experimental source: NCTC 2916

R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.

FEBS Lett. 376, 41-44, 1995

A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components

A;Reference number: S67988; MUID:96096783; PMID:8521962

A;Accession: S68220

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-12 <FLU>

A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:G2160224

R;Betley, M.J.; Somers, E.; Dasgupta, B.R.

Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989

A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term

A;Reference number: A33401; MUID:89350959; PMID:2669749

A;Accession: A33401

A;Molecule type: DNA

A;Residues: 1-35 <BET>

A;Cross-references: GB:M27892; NID:G144880; PIDN:AAA23269.1; PID:G551776

R;Gimenez, J.A.; Dasgupta, B.R.

J. Protein Chem. 12, 351-363, 1993

A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an

A;Reference number: A53884; MUID:94000342; PMID:8397793

A;Accession: A53884

A;Status: preliminary

A;Molecule type: protein

A;Residues: 867-880;1148-1217 'Y', 1219 <GIM>

A;Experimental source: strain Hall

A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex

R:DaGupta, B.R.; Dekleva, M.L.
 Biochimie 72, 661-664, 1990
 A:Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and around the active site
 A:Reference number: A60025; MUID:91120847; PMID:2126206
 A:Accession: A60025
 A:Molecule type: protein
 A:Residues: 2-6; 445-453, 'X', 455-457 <DAS1>
 R:DaGupta, B.R.; Foley, J.; Niece, R.
 Biochemistry 26, 4162, 1987
 A:Title: Partial sequence of the light chain of botulinum neurotoxin type A.
 A:Reference number: A27000
 A:Accession: A27000
 A:Molecule type: protein
 A:Residues: 2-47 <DAS2>
 R:Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.; J. Biol. Chem. 269, 1617-1620, 1994
 A:Title: Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.
 A:Reference number: A49708; MUID:94124495; PMID:8294407
 A:Contents: annotation
 C:Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synapses
 C:Genetics:
 A:Gene: atx; bota
 C:Function:
 A:Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associated protein 25
 C:Superfamily: tetanus toxin
 C:Keywords: disulfide bond; hydrolase; metalloprotease; neurotoxin; transmembrane protein
 F:2-44/Product: botulinum A light chain #status experimental <LIGHT>
 F:445-1296/Product: botulinum A heavy chain #status experimental <HY>
 F:223,227/Binding site: zinc (His) #status predicted
 F:224/Active site: Glu #status predicted

Query Match 99.2%; Score 6768; DB 1; Length 1296;
 Best Local Similarity 99.9%; Pred. No. 1.7e-287;
 Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MPFVNQPNYKDPVNGVDIAIKIPNAGQMPVKAFLHNNKIWIPIERDTPNPEEGDLN	60
Db	1	MPFVNQPNYKDPVNGVDIAIKIPNAGQMPVKAFLHNNKIWIPIERDTPNPEEGDLN	60
Qy	61	PPPEAKQPVSYDYSTYLSLTNEKNDYLGKVTLPRIYSDLDGRMLLTSLVRGIPFWGG	120
Db	61	PPPEAKQPVSYDYSTYLSLTNEKNDYLGKVTLPRIYSDLDGRMLLTSLVRGIPFWGG	120
Qy	121	STIDTELKVIDTNCINVLQDGSYRSEELNLIIGPSADIIQFECKSFGEVLNLTNGY	180
Db	121	STIDTELKVIDTNCINVLQDGSYRSEELNLIIGPSADIIQFECKSFGEVLNLTNGY	180
Qy	181	GSTQYIRFSPDFTFGFEESLEVDNPLLGAKGFATDPVTLAHLIYAGHRLYGIAPNP	240
Db	181	GSTQYIRFSPDFTFGFEESLEVDNPLLGAKGFATDPVTLAHLIYAGHRLYGIAPNP	240
Qy	241	RVFKVNTNAYYEMSGLEVSPEELRTFGGHDAKFIDSLQENFRLYYYNKPDKDIASLTNKA	300
Db	241	RVFKVNTNAYYEMSGLEVSPEELRTFGGHDAKFIDSLQENFRLYYYNKPDKDIASLTNKA	300
Qy	301	KSIVGTGTASLQVMKNVFEKYLSDTSKGFSVDKLFKDKLYKMLTEIYTDNEDNVKPKFV	360
Db	301	KSIVGTGTASLQVMKNVFEKYLSDTSKGFSVDKLFKDKLYKMLTEIYTDNEDNVKPKFV	360
Qy	361	LNRKTYLNFDAVKFINKIVPKNYTIYDGFNLNLTNLAANFGQNTNINNNPFLKNFT	420
Db	361	LNRKTYLNFDAVKFINKIVPKNYTIYDGFNLNLTNLAANFGQNTNINNNPFLKNFT	420
Qy	421	GLFEYKLLCVRGITTSKSLDKGNKALNDLCIKVNNWDLFTSPSDNFTNDLNKGE	480
Db	421	GLFEYKLLCVRGITTSKSLDKGNKALNDLCIKVNNWDLFTSPSDNFTNDLNKGE	480
Qy	481	ITSDTNIAEAEENISLDLIQYYITFTNFDNPNENISLENLSDIIQGLLELMPNTERPFG	540
Db	481	ITSDTNIAEAEENISLDLIQYYITFTNFDNPNENISLENLSDIIQGLLELMPNTERPFG	540
Qy	541	KYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVVTFPSSDYVKKVKNKATEA	600
Db	541	KYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVVTFPSSDYVKKVKNKATEA	600

Db	541	KKYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVVTFPSSDYVKKVKNKATEA	600
Qy	601	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDDFVGLIFSG	660
Db	601	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDDFVGLIFSG	660
Qy	661	AVILLEFIPEIAIPVLGTALVSIANKVLTVTOTIDNALSQRNKKWDEVVYKIVTNWLAK	720
Db	661	AVILLEFIPEIAIPVLGTALVSIANKVLTVTOTIDNALSQRNKKWDEVVYKIVTNWLAK	720
Qy	721	VNTQIDILIRKKMEALENQAEATKAIINVOYNOYTBEEKKNINFNIDDLSSKLNESINKA	780
Db	721	VNTQIDILIRKKMEALENQAEATKAIINVOYNOYTBEEKKNINFNIDDLSSKLNESINKA	780
Qy	781	MINIKFLNOCVSYLMNSMIPYGVKRLDFDASLADALIKYIYDNRGTIGQVDRDKDK	840
Db	781	MINIKFLNOCVSYLMNSMIPYGVKRLDFDASLADALIKYIYDNRGTIGQVDRDKDK	840
Qy	841	VNNTLSTDIPFQLSKYVDNQRLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI	900
Db	841	VNNTLSTDIPFQLSKYVDNQRLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI	900
Qy	901	GSKVNFDPIDKNOIQLFNLESSKIEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN	960
Db	901	GSKVNFDPIDKNOIQLFNLESSKIEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN	960
Qy	961	EYTIINCMMNSGKVSILNAGEIITWTQDQTEIKQKRVFKYSQMINISDYINRWIFVTIT	1020
Db	961	EYTIINCMMNSGKVSILNAGEIITWTQDQTEIKQKRVFKYSQMINISDYINRWIFVTIT	1020
Qy	1021	NNRLNNSKIYINGRLIDQKPISNLGNHASNIMFKLDGCRDTHRYIWIYKFLFKELN	1080
Db	1021	NNRLNNSKIYINGRLIDQKPISNLGNHASNIMFKLDGCRDTHRYIWIYKFLFKELN	1080
Qy	1081	EKEIKDYDNQNSGILKDFWGDYLDKPYVLMNLYDPNKYVDVNNVGRGYWYLGKPR	1140
Db	1081	EKEIKDYDNQNSGILKDFWGDYLDKPYVLMNLYDPNKYVDVNNVGRGYWYLGKPR	1140
Qy	1141	GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA	1200
Db	1141	GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA	1200
Qy	1201	GVKILSALBEIPDVGNLSQVVMKSKNDQGITNCKRNLQDNNGNDIGFGFHQFNNAK	1260
Db	1201	GVKILSALBEIPDVGNLSQVVMKSKNDQGITNCKRNLQDNNGNDIGFGFHQFNNAK	1260
Qy	1261	LVASNWNROIERSSRTLGCSEWFIPIVDDCWGERPL 1296	
Db	1261	LVASNWNROIERSSRTLGCSEWFIPIVDDCWGERPL 1296	

RESULT 2

140645 botulinum neurotoxin type A - Clostridium botulinum

C:Species: Clostridium botulinum

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004

C:Accession: I40645

R:Williams, A.; East, A.K.; Lawson, P.A.; Collins, M.D.

Res. Microbiol. 144, 547-556, 1993

A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A

A:Reference number: 140645; MUID:94143603; PMID:8310180

A:Accession: I40645

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1296 <RES>

A:Cross-references: UNIPROT.Q45894; EMBL.X73423; NID:G507070; PIDN:CAA51824.1; PID:G507070

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 90.3%; Score 6158; DB 2; Length 1296;

Best Local Similarity 89.9%; Pred. No. 6.4e-261;

Matches 1165; Conservative 66; Mismatches 65; Indels 0; Gaps 0;

Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTTPEEGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTTPEEGDLN 60
Qy 61 PPEAKQVPVSYDYDSTYLSSTNEDKNTLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
Db 61 PPEAKQVPVSYDYDSTYLSSTNEDKNTLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGSGYRSSEELNLIIGPSADIIQFECKSFHGVNLTNNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGSGYRSSEELNLIIGPSADIIQFECKSFHGVNLTNNGY 180
Qy 181 GSTOYIRFSDPFTGFPESLEVDNPLLGAGKATDPAVTLAHELIIYAGHRLYGIAPNP 240
Db 181 GSTOYIRFSDPFTGFPESLEVDNPLLGAGKATDPAVTLAHELIIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFTDSIQENEFRLYYYNKKPDASTLANKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFTDSIQENEFRLYYYNKKPDASTLANKA 300
Qy 301 KSIIVGTTASLQYMNKVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSIIVGTTASLQYMNKVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFQKAVFKINIYVKNVYTIYDGNLNRNLTNLAANFGONTINNNWFTKLKNFT 420
Db 361 LNRKTYLNFQKAVFKINIYVKNVYTIYDGNLNRNLTNLAANFGONTINNNWFTKLKNFT 420
Qy 421 GLPEFYKLLCVRGIIITKTSYSLDKGVNKAANDLICIKVNNWDLFSPSEDFNTDLNGEE 480
Db 421 GLPEFYKLLCVRGIIITKTSYSLDKGVNKAANDLICIKVNNWDLFSPSEDFNTDLNGEE 480
Qy 481 ITSNTNIEAAEENISLDLIQOYYLITFDNPEPENISIEENLSSDIIQGLELMPNIERPNG 540
Db 481 ITSNTNIEAAEENISLDLIQOYYLITFDNPEPENISIEENLSSDIIQGLELMPNIERPNG 540
Qy 541 KKYELDKYTHFYLRQAQEPHKGSRIALTNSVNEALLNRSRVYTFPSSDYVKKVKNKATEA 600
Db 541 KKYELDKYTHFYLRQAQEPHKGSRIALTNSVNEALLNRSRVYTFPSSDYVKKVKNKATEA 600
Qy 601 AMELGWQELVYDFTDETSVSTWDKADIITIIPIYGPALNIGNMLYKDDFVCAILFSG 660
Db 601 AMELGWQELVYDFTDETSVSTWDKADIITIIPIYGPALNIGNMLYKDDFVCAILFSG 660
Qy 661 AVILLEIPEIETAIPLVGTFAVSYIANKVLTVQIDNALSKRNEKWDVVKYIIVNNLAK 720
Db 661 AVILLEIPEIETAIPLVGTFAVSYIANKVLTVQIDNALSKRNEKWDVVKYIIVNNLAK 720
Qy 721 VNTQIDILRQKMKDALENQAEATKAIINYQVQYTEBEEKNNINFDLSSKKNESINKA 780
Db 721 VNTQIDILRQKMKDALENQAEATKAIINYQVQYTEBEEKNNINFDLSSKKNESINKA 780
Qy 781 MININKELNQCYSVYLNKSMIPYGVKLEDPDASLKDALLKIYVDRGTLIGQVDRKDK 840
Db 781 MININKELNQCYSVYLNKSMIPYGVKLEDPDASLKDALLKIYVDRGTLIGQVDRKDK 840
Qy 841 VNNLTSTIDIPOLSKYVDNORLSTFTTEYIKNIINTSILNLRVBSNLIIDLSRYASKINI 900
Db 841 VNNLTSTIDIPOLSKYVDNORLSTFTTEYIKNIINTSILNLRVBSNLIIDLSRYASKINI 900
Qy 901 GSKVNFDPIDKQNLQFLNLESSKIEVLKNAIVNMYNFPSTFWIRIPKYFNISILNN 960
Db 901 GSKVNFDPIDKQNLQFLNLESSKIEVLKNAIVNMYNFPSTFWIRIPKYFNISILNN 960
Qy 961 EYTTINCENNSGKWSLNGEIIITWLDQTEIKQVUVFKYSQMINISDVINRWIFVTIT 1020
Db 961 EYTTINCENNSGKWSLNGEIIITWLDQTEIKQVUVFKYSQMINISDVINRWIFVTIT 1020
Qy 1021 NNRNNKSIYNGRLIDQKPTSNLGNTHASNNMFKLDGCRDTHRYIWIYFNLFDKELN 1080
Db 1021 NNRNNKSIYNGRLIDQKPTSNLGNTHASNNMFKLDGCRDTHRYIWIYFNLFDKELN 1080
Qy 1081 EKEIKDLYDQNSGILKDFWGDYLOYDKPYMMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140

Db 1081 EKEIKDLYDQNSGILKDFWGDYLOYDKPYMMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
Qy 1141 GSVMTTNIYVNSLYRGTKFIIKKYASGNKDNIVRNNDRVYINNVKNKEYRLATNASQA 1200
Db 1141 GSVMTTNIYVNSLYRGTKFIIKKYASGNKDNIVRNNDRVYINNVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALEIPDVGNLSQVVMKSKVDQGITNKKCKNLQDNNNGNDIGTGFHGFNNIAK 1260
Db 1201 GVEKILSALEIPDVGNLSQVVMKSKVDQGITNKKCKNLQDNNNGNDIGTGFHGFNNIAK 1260
Qy 1261 LVASNNYNRQIERSSRTLCGSWEFIPVDDGWERPL 1296
Db 1261 LVASNNYNRQIERSSRTLCGSWEFIPVDDGWERPL 1296
RESULT 3
S33411
botulinum neurotoxin type F - Clostridium baratii
C;Species: Clostridium baratii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium baratii type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: UNIPROT:Q45851; EMBL:X68262; NID:g49138; PID:CAA48329.1; PID:g49138
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
Query Match 36.7%; Score 2504.5; DB 2; Length 1268;
Best Local Similarity 41.4%; Pred. No. 1.1e-101; Indels 107; Gaps 36;
Matches 552; Conservative 242; Mismatches 433;
Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTTPEEGDLN 59
Db 1 MPVNNFNVNDPINNTTILYMKMPYVEDSNKYKAFEMDNVWIIPERN-IIGKKPSDF 59
Qy 60 NPPEAKQVPVSYDYDSTYLSSTNEDKNTLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWG 119
Db 60 YPPISLDSGSSAYDYDENYLTDAEKDRFLKTVIKLFNRINSNPAGQVLLBEIKNGKPYLG 119
Qy 120 GSTIDTELKVIDTNCIN-----VIQPDGSGYRSSEELNLIIGPSADIIQFECKSFGEV 172
Db 120 -----NDHTAVNEFCANNRSTSVIEKESNGTTDSMLNLIIVILGPGNII--EGSTFPVRI 172
Qy 173 L-----NLTRNGYSGTQYIRFSPDFTFGPEESLEVDNPLLGAGKATDPAVTLAHELI 226
Db 173 FPNNAIYDSEKFGSGIQLMSFSTEYEAFTNDTL-----FIADPAISLAHELI 222
Qy 227 YAGHRLYGI-AINPNRVKNTNAYEMSGLEVSFEELRTFGGHDAKFTDSIQENEFRLY 285
Db 223 HVLHGLYGAQVNTKKVIEVDQCALM-AAEKDIKIEBITFGGQDLNLIITNSTNQKIYVI 281
Qy 286 YNNKFKDIASTL---NKAKSIVGTTASLQYMNKVPKFKYLLSEDTSGKFSVDKLFKDKLY 342
Db 282 LLSNYTAIASRLSQVNRNNSALNTT---YKKNPFQMKYGLDQDSNGNYTVNISKFNAYI 337
Qy 343 KMLTEIYTEDNFVKFKVNLNRKTYLNFDAKVFKNITVPKVNITYYDGNLNRNLTNLAANFN 402
Db 338 KKLFS-FTCEDLAQKQVKNRSNLYLPHFKPFRLLDLDLDDNIIYSISEGFNI--GSLRVNN 394
Qy 403 GQNTENNNMFTKLKNFTGLPEFYKLLCVRGIIITSK-TKSLDKGVNKAANDLICIKVNNW 461
Db 395 GQINILNRSIRVGPIDP-NGLVERFVGLC-KSIVSKKGTG-----NSLCTIKVNNRD 442
Qy 462 LFPSPSEDFNT-NDLNKGEBITSDTNIEAAEENISLDLIQOYYLITFNFDNPEPENISIE 520
Db 443 LFPVASESSYNGINGINSPKEI-DDTTITNNNYKKNLD---EVILDYNSD-----AIPNL 492

Qy 521 SSDIIGQL-----ELPMNIEFPNG-----KKYELDKYTMFHYLRAOBEHCKSRIALTNSV 572
Db 493 SSRLLNTAQNDSDVPKYD--SNGTSIKBYTDVQKLNVFFLYAQAPEGESALSTSSV 550
Qy 573 NEALLNRSVTFSSDYVKKVKATAAAMFLGVEQLVYDFDTESEVSTTDKADIITI 632
Db 551 NTALLDASKVTFSSDFINTVKNPQVAALFISWQQVINDFTTEATQKSTIDKIADISL 610
Qy 633 IIPYIGPALNIGMNLKYDDFVGALIFSGAVILLEFFIEPIAIPVLGTALVSYI-----AN 687
Db 611 IVPYVGLALNIGNEVQGNFKPAIEELLCAGILLFVPELLIPTILVFTIKSFNSDDSKN 670
Qy 688 KVLTVQVIDNALSRNEKQWVVKYIVTNWLAKVNTQIDLRKKKKEALENQAEATKAI 747
Db 671 KI--IKAINNALRELKWKVEYISVSNMLTRINTQFKRKEQYQALQNDQVGIKKII 728
Qy 748 NYQYNQVTEEEKNI--NFNIDDLSSKLINESINKAMINIKFLNQCVSVYLMNSMIPYGV 805
Db 729 EYKNNYTLDEKRLRAEYNIYSIKEELNKKVSLAMQNDRELFTESSISYLMKLINEAKI 788
Qy 806 KRLDFDASLKDALKYIYNRGTL-IGQVDRLLKDKVNNLTSTDPQLSKYVDNQRLLS 864
Db 789 NKLSEYDKRVNQYLLNTYLENSSTLGTSSVPELANLVSNLTNLSIPPSELYTNDKTLIH 848
Qy 865 TFFEYIKNINTSLNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIOLFLESKI 924
Db 849 ILIRFYKRIIDSSILNKKIENRRNIDSSGYCSNISINGDIYIYSTNRNQGISSRLSEV 908
Qy 925 EVILKNAIVNMSYENFSTFWIRIPKYPFNSISLNNEYTTINCM-ENNSGWKVSILNYGEI 983
Db 909 NITQNTIYNSRYQNFVSFPMVRIPKYNNLKNLNEYTTINCMNRNNSGWKISLNNNI 968
Qy 984 IWTLODQIEIKQVVFKYSQMINISDYINRWIFVTITNRLNLSKIYINGELIDOKESIN 1043
Db 969 IWTLODQTTGNOKLVNFYQMDISDYINKWTFVTITNRLHSGKLYINGNLTDQKSLN 1028
Qy 1044 LGNTHASNNIMFKDGGCDTHRYIWIYFNFLDKELNEKETKOLYDQNSGILKDFWGD 1103
Db 1029 LGNIHVDNIIIFKIVGCDNT-RYVGIRYFKIFNMELDKTEIETLYHSEPDSTILKDFGN 1087
Qy 1104 YLOYDKPYMLNLDPNKYVDNVNNGIRGYMYLKGPRGSVMNTIYNLSLYRTKFIK 1163
Db 1088 YLLYNNKYLLNLLKPNMSVTKNS-----DIILNINRQGIYSKTIWIFSNARLYTGEVIR 1143
Qy 1164 KYAS---CNKNNIVRNDRVYINVVKNKEYRLATNASQAGVEKILSALEIPDVG-NLSQ 1219
Db 1144 KVGSTDTNTDNFVRKNDTVYINVDGNSYQLYADVSTSAVEKTIKLRISNSYNSNQ 1203
Qy 1220 VVWKSNDQGITNKKCNLODNGNDIGFIGHQFNIAKLVASNWTNRQIERSRTL 1279
Db 1204 MIIMDS-----IGDCTWNFKTNGNDIGLGFH-LNN--LVASSWYKNNRNRNG 1254
Qy 1280 CSWEPFVDDGWE 1293
Db 1255 CFWSFISKEHGWE 1268
RESULT 4
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C:Species: Clostridium botulinum
C>Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 09-Jul-2004
C:Accession: S21178; S48107; JH0257; B35294; A60027; S1811
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A:Reference number: S21178; MUID:92174922; PMID:1541280
A:Accession: S21178
A:Molecule type: DNA
A:Residues: 1-1252 <WHE>
A:Cross-references: UNIPROT:Q00496; UNIPROT:Q45862; EMBL:X62683; NID:G40397; PIDN:CAA445
R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A:Reference number: S48103; MUID:94013372; PMID:8408542
A:Accession: S48107
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 616-982 <CAM>
A:Cross-references: EMBL:X70815; NID:G407786; PIDN:CAA50146.1; PID:G407787
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type
A:Reference number: JH0256; MUID:92181428; PMID:1543481
A:Accession: JH0257
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-176, 'R', 178-197, 'C', 199-339, 'R', 341-772, 'I', 774-962, 'FE', 965-966, 'R', 968-1
A:Cross-references: EMBL:X62089; NID:G40393; PIDN:CAA43999.1; PID:G40394
R:Binz, T.; Kurazono, H.; Willie, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A:Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A:Reference number: A35294; MUID:90264400; PMID:2160960
A:Accession: B35294
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-176, 'R', 178-252 <BIN>
A:Experimental source: strain Beluga
R:Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A:Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A:Reference number: A60027; MUID:90344918; PMID:2116911
A:Accession: A60027
A:Molecule type: protein
A:Residues: 420-427 <GIN>
A:Experimental source: strain Beluga
A:Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C:Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin
F:1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F:423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F:412-426/Disulfide bonds: #status predicted
Query Match 34.3%; Score 2339; DB 2; Length 1252;
Best Local Similarity 39.9%; Pred. No. 1.8e-94;
Matches 535; Conservative 249; Mismatches 421; Indels 136; Gaps 43;
Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGQNPVKAFKHNKIWIPIERDTF-TNPEEGDL 59
Db 1 MPKIN-SFYNDPVDNRITLYIK--PGGCOEFYKSFNIMKNIWIPIERNVIGTFPQ--DF 55
Qy 60 NPPEAKQVPSVYDSTYSLTDNEKNYLGVTKLPERIYSTDLGRMLTYSIVRGIPFWG 119
Db 56 HPPTSLKNGDSSYDYPNYLOSDEKDRFLKIVTKIFNINNNLSSGILLBELSKANPYLG 115
Qy 120 G-STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIOFECKSGHEVLN---L 175
Db 116 NDNTPDNQFHIGDASAVEIKFNSGSDILLPNVIINGAEPDL--FETNSSNISLRNMP 173
Qy 176 TRNGYGTQVIRSPDFTFGFESLEVDTPNLLGAGKFAFDPAVLAHELIVAGHRLV-- 233
Db 174 SNHGFSGIAIVTFSPEYSFRNDN-----SMNEFIQDPALTLHMLHLSLHGLYGA 224
Qy 234 -GIAINERNVFKVN---TNAYEMSGLEVSFEELRTFGGHDAKFDLSLOENFEFLYYNK 289
Db 225 KGITTKTYIQKNPLTN-----IRGTNI--EEFLTGGTDLNIIITSAQNDIVTNLLAD 278
Qy 290 FKDIASLTNKAISVGTGTTASLQYMNKVFKEKYLSEDTSGKFSVDKLPKPKLYKMLTEIY 349
Db 279 YKTIASKLSKQV---SNPLINPYKDYFEAKYGLDKDASGIYSVYNKFNKFNDFPKGLYS-F 334

Db 867 DSSILDMRYENKPFIDISGYSNLSINGNYIYSTNRNQFGIYNLSRSLSEVNIQAQNNDIY 926
Qy 935 NSMYENFSTFWIRIPKYFNISLNNEYTIINCH-ENNSGWKYSLN-GEIITWLODT 990
Db 927 NSRYQFISFWIRIPKHYKPMHNRREYTIINCNNNSGWKISLRVDRCEIITWLODT 986
Qy 991 QETKQVVKYQOMINSDYINRWIFVTITNNRLNNSKIYINGRLIDOKPISNLGNTHAS 1050
Db 987 SGNKENLIPRYEELNRISYNKIFVTITNNRLNNSRIYINGNLIVEKISNLGDHVS 1046
Qy 1051 NNIMFKLDGCDTHRYIWKYFNLPDKELNEKEIKDYDNQSNSTIILKDFWGYLOYDXP 1110
Db 1047 DNILFKIVGC-DBETYGIRYFKVFNTKTEIETLYSENEPDSILKNYWGNYLLYKX 1105
Qy 1111 YMLNLDPNKYVDVNVGIRGYMYLKGPRGSMVTNIVLNSSLYRGTKFIKKYAS-- 1167
Db 1106 YLFLNLKDKYITLNS---GILNINQQRG-VTEGVSFLNYKLYEGVEIIRKNGPIDI 1160
Qy 1168 GNKDNIVRNRDRYINVVVKNKEVRLATNASQAGVEKILSALEIPDVGNLSQVVMKSKN 1227
Db 1161 SNTDNFVRKNDLAYINVDVGRVRYLAD-TKSEKEIIRTSNLND--SLGQIIVMDS-- 1215
Qy 1228 DOGITNKKCNLQDNNGNDIGFGFHPFNIAKIIVASWYNNRQIERSRSLGCSWEPIV 1287
Db 1216 ---IGNNCTMNFQNNNGSLGLGFHSNN---LVASSWYNNIRRTSSNGCFWSSISK 1268
Qy 1288 DQGWGE 1293
Db 1269 ENGWKE 1274

RESULT 7
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:9441275; PIDN:CAA52275.1; PID:9441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 33.5%; Score 2285; DB 2; Length 1297;
Best Local Similarity 39.6%; Pred. No. 4.2e-92;
Matches 531; Conservative 217; Mismatches 502; Indels 90; Gaps 29;

Qy 1 MPFVNKQFNKDPVNGVDIAIYK-IPNAGQMPKAFKHNKIWIPIERDTF-TNPEEGD 58
Db 1 MPVNIKNFNNDPINDDIIMPEFPNDPGPGTYKAPRIIDRIWIVPERFYGPQDPN 60
Qy 59 LNPPEAKQPVSVYDSTYSLTDNEKDNLYKGVTKLPERIYSTDLGERMLTSLVRGIPFW 118
Db 61 ASTGVFSKDV-YEYIDPTYLTKDAEKDKFLTKMILFNRSKPSGQRLDDMIIVDAIPLY 119
Qy 119 GGSTIDPELKVDTNCLNV-----IQPDGSYRSEEL--NLVITGPS---ADIIQPECKS 167
Db 120 GNAS--TPPKFAANVANVINKKIIQPGAEQIKGLMTWLLIIFGPGVLSDFNFTDSMIM 177
Qy 168 FGHEVLNLRNGYSGTQYIRFSPDFTFGFESLEVDTPNLLGAGKFPATDPAVTLAHELIY 227
Db 178 NGHSPIS--EGFGARMIRFCPSCLNVNQNENKDTISFRRAYFADPALTLAHELIH 234
Qy 228 AGHRLYGIANPNRVKNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRIYY 287
Db 235 VLHGLYGIKIS-NLPITPTNKEFFMQHSDPVOABEITYFGCHDPSVISPSTD----MNIY 289

Qy 288 NK----FKDIASLTNKAKSIVGTATSLQYMNKVNFKKYLSEDTSGKFSVDKLFKDKLYK 343
Db 290 NKALQNFODIANRLNIVSSAQSGIDISLYKQIYKQYDFVEDPNGKYSVDKDFKDKLYK 349
Qy 344 MLTEIYEDNFVFKFVKLARKTYLNFDKAVFKINIVPKVNYTYI---DGNLRLNTNLAAN 400
Db 350 ALMFGFTETTLAGEYGIKTRYSY--FSEYLPPIKTEKLDNTIYTONEGENIASKNLUTE 407
Qy 401 FNCQNTENNNMFTKLNFTGLPEFVKLCVGRGIITSKTSKLDKGVNKALNDLICIKVNNW 460
Db 408 FNCQKAVNKEAYEES--LEHLVIVRIAMCKPVMYK-----NTGSEQCIIYVNE 456
Qy 461 DLFPFSEDNFTNDLNGEISITSDTNEAABENISLD-LIQYYLTFNFNEPENEISIEIN 519
Db 457 DLFFIANKDSFKDLAKAETIAYNTQNTIENNFSIDQLILDNDLSSGIDPLNENTPEFT 516
Qy 520 LSSDI-----IGOLELMPNIERPNGKKYELDKYTHPHYLRAQEFPHGKGRALITSVNE 574
Db 517 NFDDIDIPVYIKOSAL-----KKIPVDGSLFEYLHAQTFPSNIENLQJLTSNLD 566
Qy 575 ALLENPSRVYTFSSDVVVKVKNKATEAAMFLGWGEVQLVYDFDTSETSEVSTTDKIADIRII 634
Db 567 ALRNNKVTYTFSTNLVEKANTVVGASLFWVWVGVIDDFESTQKSTIDKYSDVSIII 626
Qy 635 PYIGPALNIGNMLYKDDFVGALIFSGAVILLEPIEPIAIPVLGTFALVSYIANKVLTVQT 694
Db 627 PYIGPALNVGNETAKENFKNAPEIGGAAILMEFIPELIVPIVGFTLESYVGNKGHIIMT 686
Qy 695 IDNALSKRNEKDEVVYKIVTNWLAKVNTQIDILRKKMKEALENOABATKALINYQNOY 754
Db 687 ISNALKKRQKQWTDMYLVSQWLSVTNTQFYTKERMYNALNNQSAIEKIIDQYNR 746
Qy 755 TEEKNNINFINIDLSKLNESINKAMINIKFNLNCSVSYLMNSMIPYGVKLEDFDAS 814
Db 747 SEDKNNINIDFNDIDFKLNQSNLAINNIDDFINQCSISYLMNRMIPVAKKLKDFDN 806
Qy 815 LKDALIKYIYDNRGTILIGQVDRKDKVNNLTSTDIPQLSKYVDNQRLSLTFTEYIKNII 874
Db 807 LKRDLEIYDITNELYLLDEWILKSKVNRHLKDSIPFDSLTKDTILIVFNFNYSNIS 866
Qy 875 NTSILNLRVESNHLIDLRYAKINIGSKVNPDPIDKQIOLNLESSKTEVILKNAIVY 934
Db 867 SNAILSLSYRGRLIDSSGATMNYGSDVIFNDIGNQPKLNNSENSTIAHQSKFVY 926
Qy 935 NSMYENFSTFWIRIPKYFNISLNNEYTIINCMNNNSGWKYSLNVEIITWLODTQ 991
Db 927 DSWDFNFSINFWRTPKYNNNDIQTLYONEYTIISCIKNDSGWKVSIKGRNIIWTLIDVN 986
Qy 992 EIKQVVFYKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDOKPISNLGNTHAS 1051
Db 987 AKSKSIFFEYSIKDNISDYINKWFSITITNDRLGANANIYINGSLKSEKILNDRINSSN 1046
Qy 1052 NIMFKLDGCDTHRYIWKYFNLPDKELNEKEIKDYDNQSNSTIILKDFWGYLOYDXP 1111
Db 1047 DIDFKLINCTDTTKFVWIKDFNIFGBELNATEVSSLYWIOSSTNTLKDFWGNPLRYDTQY 1106
Qy 1112 YMLNLDPNKYVDVNVGIRGYMYLKGPRGSMVTNIVLNSSLYRGTKFIKKYASG-- 1168
Db 1107 YLFNQGMQNIYIKFS---KASMGETAPRTNF--NNAAINYQNLGLRLIIFKASNRN 1161
Qy 1169 -NKDNIVRNRDRYINV-VYKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVMKSKN 1226
Db 1162 INNDNIVREGDIYILNIDNISDESRYVVLVNSKEIQTLFLAPINDPFTFYDLQIK-K 1220
Qy 1227 NDOGITNKKCNLQDNNGNDIGFIFGHPQ-----NNIAKLVASWYNNRQIER-- 1273
Db 1221 YYEKTYNCQI-LCEKDTKTFGLFGIKFKYQYGYWMDTYDNY--FCISOWYLLRRISENI 1277
Qy 1274 SSRTLGCSWEFIPVDDGWGE 1293
Db 1278 NKURLGCNMQFIPVDDGWGE 1297

QY 452 ---DLICIKVNNWDLFFSPSEDNFTNDLNGEETISDTNIEAAEENISLDLIQQVYLTFN 507
Db 461 DLGGELCIKIKNEBTLFIAEKNEFSSEBPFQDEIVSYNTKNKPLNFNYSLDKIIVDY---- 516
QY 508 FDNFENISIELSSDIQGLELMPNLEREPNGKK-----YELDKYTM 550
Db 517 -----NLQSKI-----TLPNDRTTPVTKGIPYAPBYKSNAASTIBIHINDNTI 560
QY 551 FHYLRAQEPHGKSRIALTNSVNEALNPSRVYTFSSDYVKYKNAKATEAAAMFLGWVQL 610
Db 561 YOYLYAQSPTTLQRTIMTNSVDDALINSKIYSYFSP-VISKVNOQAQGLFLQWVRDI 619
QY 611 VYDFTDETSEVSTDKIADITIIPYIGPALNIGMLYKODFVGALIFSGAVILLEFPIE 670
Db 620 IDDFTESSQKTTIDKISDSTVTPYIGPALNIVKQYEGNFIGALETTGVVLLLEYPIE 679
QY 671 IAIPLVGLTALVSYIAKVLTQOTIDNALSKRNEKWDVEYKIVTVMIAKNTQIDLRK 730
Db 680 ITPVTAUASIAESSTQKEIKIITDINFLEKRYEKVIEYKLVAKWLGTVNTQPKRSY 739
QY 731 KMKALENAEATKAIINYOYNOYTEBEKNINENIDOLSKNESINKAMINIKFLNQ 790
Db 740 QMYRSLEYQVDAIKKIIDYIYKISGPDKEQIADENLNKNLEKANKAMINIFWE 799
QY 791 CSVSYLMNSMIPGVGRLEDFDASLDKALLKIYVDNRGTL-IGQVDRLKDKVNNLTSTDI 849
Db 800 SSRSFLVQMINAEAKQLLEFDQTSKNILMQYIKANSKFIGITELKKLESKINKVFSTPI 859
QY 850 PFQLSK-----YVDNQRLSTFTYEIKNIINTSILNLYESNHLIDLSYASKINIGSKVN 905
Db 860 PFSYKNLDCWVNEEDIDVLK-----KSTILNLDINNDIISDISGFNSVITYPDAQ 913
QY 906 FDP-IDKNOIQLFNLSSKIEVILKNAIVNSMYENFSTFWIRIPKYFNISLN----- 959
Db 914 LYPGINKKAIHLVNNESSEVIVHKAMDIEYNDMENFTVSPWLVPK-----VSASHLEQY 969
QY 960 --NEYTIINCMMN-----SGWKVSLNYGELIWTLDQTEIKQVVPKYSQMINISD--- 1009
Db 970 GTNEYSIISSMKKHSLSIGSGWSVSLKGNLITWLKDSAGEVRQITFR-----DLPDKFN 1024
QY 1010 -YI-NRWIFVTITNNLNNSKIYINGELIDQKPISNLCNTHASNNIMEKLDGCRDTHYI 1067
Db 1025 AYLANKKWFTITNDRSLSSANLYINGVLMGSAEITGLGAIBEDNNITLKLDRCNNNQIV 1084
QY 1068 WTKYFNLDPKELNEKEIKOILDQNSGILKDFMGDYLQDKPYMYMLNLYDPNKYVDVNN 1127
Db 1085 SIDKFRIFCKALNPKEIKLYTSVLSITFLDFWGNPLRYDTEYLYIPVASSKDVQLKN 1144
QY 1128 VGIrgYMYL-KGRGVSMTTNIYLNSSLYRGTKPIIKKYASGNK-DNIVRNDRVYINVV 1185
Db 1145 --ITDYMILTNPASyTNGKLNIIYR-RLYNGLKEIKRYTTPNNEIDSPVKSDFIKLYVS 1201
QY 1186 VKNKEY-----RLATNASQAGVEKILSALEIPDVGNLSQVVMVMSKN 1227
Db 1202 YNNNEHIVGYPKDGNAFNNILDRILRVGNAPGIFLYKMEAVKURDLK----- 1249
QY 1228 DQGITNKKCNLQDNNNGNDIGFGEH--QFNNAIK---LVASNNWYNOIERSRRTLGCSW 1282
Db 1250 -----TVSQVLKLYDDKVASGLGVGTNGQIGNDPNRDILIASNWFYHNLK--DKILGCDW 1303
QY 1283 EPIPVDDGW 1291
Db 1304 YFVPTDEGW 1312

RESULT 11
S11455
botulinum neurotoxin type D - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S11455
R:Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegl
Nucleic Acids Res. 18, 5556, 1990

A:Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type A
A:Reference number: S11455; MUID:91016853; PMID:216736
A:Accession: S11455
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Molecule type: DNA
A:Residues: 1-1276 <BIN>
A:Cross-references: UNIPROT:P19321; EMBL:X54254; NID:G40395; PIDN:CAA38175.1; PID:G40396
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match	27.7%	Score	1890.5	DB 2	Length	1276
Best Local Similarity	33.1%	Pred.	No. 6.4e-75			
Matches	440	Conservative	260	Mismatches	530	Indels
			99	Gaps	31	
Qy	6	KQENYKDPVNGVDIAVIKIP-NAGQQVPVKAFAKHNIWIPER-DTFTNPEEGDLNPPP	63			
Db	6	KQFNISDPVNDNDIILYLRIPQNKLTTPVKAFAITQNIWIPERFSSDTPN---SLSKPP	62			
Qy	64	EAKQVPVSYDSTYLSTDNKKNYLKGVTKLFRYIYSTDLGRMLLTISVIRGIPFWGGSTI	123			
Db	63	RPTSKYQSYDPSYLSSTQKDTFLGKIILFKRINERDYGKGLINYLWVSGPFGDST	122			
Qy	124	DTCLKVIDTNCINVTQ---PDGYSRSEEL---NLVIGPSADIIQPECK-SFGHEVLNLT	176			
Db	123	PEDTFDRHTHTNIAIAVEKFPENGSGKVTNIITPSVLIFGPLENIDYTASLTQOQSSNP	182			
Qy	177	RNGYSTQVIRFSPDFTFGFEESLEVDTPLLCAGKGFATDPVTLAHELIYAGHRLYGIA	236			
Db	183	FEFGTGLSILKVPAPFLLTSDVTSNQSSAVLGKSIPCMOPVIALMHETHSLHQLYGIN	242			
Qy	237	I-NPNRVFKVNTNAYVEMSGLEVSFEELRTFGGHDADFIDSLQENBFRLYYNNKFOIAS	295			
Db	243	IPSDKRIIRPOVSEGFSQDGPVQFELYTFGGLDVEIIPQIERSQLREKALGHYKDI	302			
Qy	296	TLNKAKSIVGTT--ASLQVMKVNFKKYLISEDTSGKFSVDKLPKDKLYKMLTELYT	353			
Db	303	RLNNINKTIPSSWISNIDYKIKIFSEKYNFDPKDTGTFVNVNIDKFNLSYDLTNMSEV	362			
Qy	354	FVKFFKVLNRKTYLNFDPKAVFKINIVPKVNYTYIDGNLRNTNLAAFNPGQNTENNMF	413			
Db	363	YSSQNVNKNRTHYFSRHYLPVFANILDDNIYTRDGNLTNKGNIENSGQNI	421			
Qy	414	TKLKNFTGLFEFYKLLCVRGITSTKSLDKGYNKALNDLCIKVNNNDLFFSPSDNFTN	473			
Db	422	QKLSS-ESVVDLTKVCLR--LTNKS RD-----DSTCIKVKNNRLPYVADKDSISQ	469			
Qy	474	DLNKGEEITSOTNIEAAENISLDLQOYLYTNFDNEPENISIENTSSDIIGQLELMPN	533			
Db	470	BIFENKIITDNTVQNSDKFSLD---ESILDGQVPINPEVD-----PLP	514			
Qy	534	IERFP-----NGKKYELDKYTMFHYLRQAQEFHGKSRIALTNSVNEALLNPSR	581			
Db	515	VNMEPLNLPGEISIVFDDITKY--VDYLYNSYYLESQKLSNNVENITLTSVEEALGY	573			
Qy	582	VYTFPSSDVKVKNKATEAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPAL	641			
Db	574	IYTFPLS-LAEKVNKGQVAGLFLNWAENEVEDFTTNMKDITLTKISDVSVIPIYIGPAL	632			
Qy	642	NIGNMLYKDDFVGALIFSGAVILLEPIPEIAIPVLGTFALVSYTANKVLTVQTDNALSK	701			
Db	633	NIGNSALRGNFQAFAAGVAFLLGEGPEPTIPALGVFTFYSSIQEREKIIKTIENCLEQ	692			
Qy	702	RNEKWDEVYKIVTWNIAKVNQTQIDILIRKKMKEALENQAEATKAIINVQVNOYTEEKN	761			
Db	693	RVKRWKDSYQWVSNWLSRITTFQFNHINYOQMSLSYQADAIRAKADILEYKKSQDKN	752			
Qy	762	INFNIDDLSSKLNESINKAMINIKFNLCQSVLYMNSMIPYGVKRLDFDASLKDALL	821			
Db	753	IKSQVENLKNSLDKVISEAMNNINKIRECSVYLFKQMLPKVIDELNKFDLRTKTELIN	812			
Qy	822	YIYDRGTLIGQVDRKDKVNNTLSTDIPFQLSKYVDNQRLLSFTFEYIKMIINTSI	881			
Db	813	LIDSNNIILVGEVDRLKAKVNESFENTMPPFIETYNNSLLKDIINEYFNSINDSKLSL	872			

Db 1207 GSFIQPCNTFDYVYASQLFSSNATNRLGLSLGSYSF-----KLGDDYWFNHEYLIPIVI 1262
Qy 1271 -IERSSRTL---GCSWEFIPVDD 1289
Db 1263 KIEHYASLLSTSTHWFVPASE 1285

RESULT 13
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C:Species: Clostridium botulinum
C:Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C:Accession: S11291; A53396; S22166; A49777
R:Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A:Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A:Reference number: S11291; MUID:90370487; PMID:2204031
A:Accession: S11291
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-84, 'p', 86-1291 <H>
A:Cross-references: UNIPROT:Q93HT3; EMBL:X53751; NID:gl4905; PIDN:CAA37780.1; PID:gl4906
R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.;
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A:Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxin
A:Reference number: A53396; MUID:91024998; PMID:2222445
A:Accession: A53396
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-669, R', 671-1291 <TS1>
R:Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A:Description: Nucleotide sequence of the gene for one of the components of hemagglutinin
A:Reference number: S22166
A:Accession: S22166
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1291 <TS2>
A:Cross-references: EMBL:X62389; NID:9558175; PIDN:CAA44263.1; PID:g40390
R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A:Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A:Reference number: A49777; MUID:91282468; PMID:2059039
A:Accession: A49777
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-607 <TS3>
A:Cross-references: GB:D90210
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match 26.08; Score 1771; DB 2; Length 1291;
Best Local Similarity 32.88; Pred. No. 1.1e-69;
Matches 443; Conservative 244; Mismatches 542; Indels 122; Gaps 34;

Qy 1 MPFVKQFNQKDPVNGVDIAYIKIP-NAGQVQPVKAFKIHNKIWIPIERDTFTNPESGDL 59
Db 1 MPITINNPNYSDPVNKNILYDLTHLTLANEPEKAFITGNWVIFDR--FSRNSNPNL 58

Qy 60 NPPEAKQVPVSYDSTYSLTDNEKNYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWG 119
Db 59 NKPRVTSPPKSGYVDPNLTSDSKDTFLKEIIKLFKRINSREIGBELIYRLSTDIPFG 118

Qy 120 G-----STIDELKVIDTNCINVIQPDGS--YRSEELN--LVIGPSADIIQPECKSFCH 170
Db 119 NNNTPIPTFDPD---VDFNSVDVKTROGNWVKTGSINPFSVIITGPRENIIDPETSF-- 173

Qy 171 EVNLNTRN-----GYGSTQVIRFSPDFTFGFEESLEVDNPLLGAKFA-----TDPVAV 219
Db 174 ---KLTNNTFAAQEGFGALSIIISIPRLTYSNA-----TND-VGEGRFSKSEFCMDPIL 225

Qy 220 TLAHELIYAGHRLYGIAR-NPNRVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Db 226 ILMHLELHAWHNNLYGAIAPNDQTISSVTSNIFYQYNNVKLEYAEIYAFGPTTDLIPKSA 285
Qy 279 ENEFRLLYYNKKFKDIASTLNKAISVGTATSL-----QYMKNVFKEXYLLSEDTSGK 330
Db 286 RKFPEEKALDYRSIAKRLNSI-----TIANPSSFNKYIGEYKQKLIR-KYRFVWESSGE 339
Qy 331 FSVDKLKFDKLYKMLTEIYTDNFFVKFKVNLNRTYLNFDKAVFKINIVPKVNTIYDGF 390
Db 340 VTVNRNKFVELYNELTQIFTEFNAYKIYNQNRKIYLSNVYTPVTANILDDNVDYIQNGF 399
Qy 391 NLRTNLAANFNCGNTEINNMNFKLKNFTGLFEFVKLLCVRGIIITSKTSKLDGYNKAL 450
Db 400 NIPKSNLNLVFMQNLUS-RNPALRKNVPENMLYLTK-FCHKAI---DGRSL---YNKTL 451
Qy 451 NDLCIKVNNWDLFPSPSEONFTNDLAKGBEITSDTNIEAAEENISLDLQIYYLTFNFDN 510
Db 452 DCRELLAVKNTDLPFIGDISDVKTDFLRKDIINEETEVIYYPDNVSD--QVILS---KN 505
Qy 511 EPNISIEENSSDIIIGOLELMPNIR-FPNKKYKELDKYTMFHYLRAQEFEGHKSRIALT 569
Db 506 TSEHGQLDLLYPSIDSSESEILPGENQVFDNRQTQNDYLSNYLYLESQKLSDNVEDFTFT 565
Qy 570 NSVNEALLNPSRVYTFESSDYVKKVNAKATEAAMFLGWVEOLVYDFDTEDETSEVSTDKIAD 629
Db 566 RSEEEALDNSAKVYTFPT-LANKVAGVQGGFLMANDVVEDFTNIIIRKDTLTKISD 624
Qy 630 ITTIIYVIGPALNIGNLYKDDFVGALIFSGAVILLEFIEPIAIPVLGTALVSYIANKV 689
Db 625 VSAIYIIGPALNISNSVRGNTEAFVGTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
Qy 690 LTVQTDNALSKNEKWDVEYKIVTNLWAKVNTQIDLIRKKMKEALENAEATKALINY 749
Db 685 EIIKTIIDNCLEQRIKWKDSYEMWMTLSRIITQFNNISYQYMSLNYQAGAKAKIDL 744
Qy 750 QYNQYTEEEKNNINFINDDLSSKLNKINKFLNQCSVYLMNSMIPGVKRL 809
Db 745 EYKYSGSKENIKSQVENLKNLSDVKISEAMNINKFIRECSVTYLFKMLPKVIDELN 804
Qy 810 DFDASLKDALLKYYINRGTGLIQVDRDKQVNTLSTOIPFQLSKYVDNRQLISTTEY 869
Db 805 EFDRTYAKLINLIDSHNILLVEVDKLAKVNSFQNTIPFNIFSNTNLSLLKDIINEY 864
Qy 870 IKNIINTSLNLRYSENHLIDLRYASKINIGSKVNFDPIDKNOIQLFNLESS-----KI 924
Db 865 FNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGVQLNPIFPFD---FKLGSSGEDRGV 921
Qy 925 EVILKNAIVYNSMYENFSTSFWRIRPKYFNSISLNNEYTIINCMEANSKVKVSLNYGEII 984
Db 922 IVTQENINIVYNSMYESFSISFWIRINKVWSNLP---GYTIIDSVKVNSGWSIGIISNFW 978
Qy 985 WTLQDTQEIQRVVFVKYSQMINISDYINRWIFVTITNRLNNSKIYINGRLIDQKPISNL 1044
Db 979 FTLLQNEDEQSINFSYDINSNAPGY-NKQFFVTVNNMGNMKIYINGRLIDTKVKEL 1037
Qy 1045 GNTHASNINFKLDGCRDT-----HRYIWKYFNLPKELNEKEIKDYDNQNSGI 1096
Db 1038 TGINFSKTITFEINKIPDGTGLITSDSNINMWIRDFYIFAKELDGKIDINILFNSLOQTV 1097
Qy 1097 LKDFWGDYLDYKPYVMALYDKNYVDVNNVGIKGYWYLYKGPBGSVWTTNIYLSLSYR 1156
Db 1098 VKDYWGNDLRYNKEYVMNIDYLNRYMANS-----RQIVFNTR-RNNNDPNE 1144
Qy 1157 GTKFIITKCKVAGSKNDIVRNDRVYINVVKNKEYRL-----ATNASQAGVEKILSA 1208
Db 1145 GYKILIKRIGTNTDTRVGRGDILYFDMTINNKAYNLFMKNETMYADNHTEDIYALGLR 1204
Qy 1209 LEIPDVGNLSQVVMKSKNDQGITNKC-KNNLODNNNGNDIGFIFGHOF-----NNI 1258
Db 1205 EOTKDINDNIIFQIQPMNNTYYVASQIFKSNFNGENISGICSIGTVYRFLGGDWYRHYL 1264
Qy 1259 AKLVASNWNRYQIERSRSLTGCSEWETPVD 1289

Db 1265 VPTVKGQNYASLLESTs-----THWGFVPVSE 1291

RESULT 14

S46431

botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)

N:Alternate names: BoNT/C1 protein

C:Species: Clostridium botulinum phage 1C

A:Variety: strain C 468

C>Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004

C:Accession: S46431; S49107

R:Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.

Mol. Gen. Genet. 243, 631-640, 1994

A:Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic P

A:Reference number: S46426; MUID:94301293; PMID:8028579

A:Accession: S46431

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1291 <HAU>

A:Cross-references: UNIPROT:Q93HT3; EMBL:X72793; NID:G516171; PIDN:CAA51313.1; PID:G5161

A:Experimental source: strain C 468

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Superfamily: tetanus toxin

Query Match 26.0%; Score 1771; DB 2; Length 1291;

Best Local Similarity 32.8%; Pred. No. 1.1e-69;

Matches 443; Conservative 244; Mismatches 542; Indels 122; Gaps 34;

Qy 1 MPFVKQFNKDPVNGVDIAIKIP-NAGOMQPVKPKHINKIWIPIERTFTNPEGDL 59

Db 1 MPITINNFNYSDFVDNKNILYDLTHLNTLANEPEKAFRITGNWIWIPDR--FSRNSPNL 58

Qy 60 NPPEAKQPVSVYDSTVLTSDNKNVGLKGVTKLPERIVSTDLGRMLTSIVRGIPFWG 119

Db 59 NKPRVTSKSGYVDPNLTSDSKOTFLKIKLFRINSREIGELIYRLSTDIPFG 118

Qy 120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIGPSADIQFECKSGH 170

Db 119 NNNTPIITFDPD---VDFNSVDVKTROGNWVKTGSINPSVIITGPRENIIDPETSTF-- 173

Qy 171 EVMLNTRN-----GYSGTVIRSPDFTGFPESLEVDNPLLGAKFA-----TDPV 219

Db 174 ---KLTNNTFAAQEGFALSIIISPRFMYLSNA---TND-VGEGRFSKSEFCMDPIL 225

Qy 220 TLAHEIYAGRLYGIAT-NPNRVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Db 226 ILMHELNHAMNLVGLIAPNDQTISSVTSNIFYQYVNVKLEYAIIYAFGGPTIDLI PKSA 285

Qy 279 ENERLYYNNKFDIASTLNKAKSIVGTASL-----QYMKNVFKKILLSSETSGK 330

Db 286 RKYFEERKALDYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVSSE 339

Qy 331 FSVKLFKEDKLYKMLTEIYEDNFVKFPKVLNRKTYLNFDKAVPKINIVKVNVTIYDGF 390

Db 340 VTNRNKKEVELYNELTQIFTEFNKIVYVQNRKILYSNVVTPVTANILDDNVVDIQNGF 399

Qy 391 NLRNTNLAANFNGQTEINNNNFTKLNFTGLFFEFYKLLCVRGHITSKTSLSLKGYNKAL 450

Db 400 NIPKSNLVLPFGQNLN-RNPAIRKVNPEENMLYFTK-FCHKAI---DGRSL---YNKTL 451

Qy 451 NDLCKVANNWDLFPSPSEDNFTNDLKGEEITSDTNEAAEENISLDLIQYIYTFNPDN 510

Db 452 DCRELLVKNLTLPFTGDISDKVTDIFLRKDNEETEVIYYPDNVSD---QVILS---KN 505

Qy 511 EPNISIEINSSDIIGOLELMPNIEP-FPNGKVKYELDKVTMFWYLAQEPFHGSKRIALT 569

Db 506 TSEHQDOLLVPSIDSESEILIPGENQVDFYDRTQNVVDVNLNYYLESQKLSDNVEDFTFT 565

Qy 570 NSVNEALLNPGRVYTFSSDYVKKYKATEAAMFLGWVEQLVYDFTDTSVSTTDKIAD 629

Db 566 RSIEEALNSAKVYTYFT-LANKVNAVQGGFLFWANDVVEDFTTNILRKDTLKDLS 624

Qy 630 ITIIPYIGPALNIGMLYKDFVICALIFSGAVILLEPIETAIPVLGTFALVSVIANKV 689

Db 625 VSAIPIYIGPALNIGNSVRGNFTFAVGTGVTILLEAPPEFTIPALGAFVIYSKVQERN 684

Qy 690 LTVQITDIALSKRNEKWDDEVYKVIYVNNLAKVNTQIDLRKKKKEALENQAEATKAIINY 749

Db 685 EIKITDNCLEORIKRWKDSYEMMGWTLRLITQFNNISYQMYDSLNYQAGAKAKIDL 744

Qy 750 QYNQYTEEFKNINFINIDLLSSKLINESINKAMININKFNQCSVSLNMSMIPYGVKRL 809

Db 745 EYKYSKSGDENIKSQVENLKNLSDVKISEAMNINKFIRECSVTYVLPKNNLPKVIDEIN 804

Qy 810 DFDASKDALLKYIYDNRGTLIGOVDRKDKVNTLSTDIPQOLSXYVDNQRLSTFTFY 869

Db 805 EFDRTKAKLILINDSHNLIILGVEVDRKAKVNSFNQNTIPFNIFSYTNNLSLKDILNEY 864

Qy 870 IKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQOLFNLSS-----KI 924

Db 865 FNNINDSKLSLQNRKNTLVDISYNAEVEEGDQLNPIFPD---FKLGSSGEDRGV 921

Qy 925 EVILKNAIYVNGMYENFSTFWIRIPKYPNSISLNNEYTIINCMMNNSGKVSLSNYGEII 984

Db 922 IVTQENIIVVNSWYSESFSFWIRINKWVSNLP--GYTIIDSVKNNSGWSIGIISNPLV 978

Qy 985 WTLQDTQEIQRVVFVKYQMINISDYINERWIFVTITNNRLANSKIYINGELIDOKPISNL 1044

Db 979 FTLKQNEDEQSFNSIDISNNAPGY-NKWFVFTVNNMGNWKIYINGKLIDITIKVKEL 1037

Qy 1045 GNIHASNIMFKLDGCRDT-----HRYIWIKYENFLFDKELNEKEIKDLYDNQSNNGI 1096

Db 1038 TGINFSEKTIITFEINKIPDGLITSDSDINNMWIRDYFPAKELDGKIDILFNLSQYTNV 1097

Qy 1097 LKDFWGDYIYQDKPYMLNLYDPNKTVDVNVVGIRGMYLKGPRGSVMTNLYNLSLYR 1156

Db 1098 VKDYGNGLRYNKYEVYVNIIDYLRNMYANS-----RQIVFNTR-RNNNDENE 1144

Qy 1157 GTRFIIKKYASGNKDNIVRNRDVIYVNVVKNKEYRL-----ATNASQAQVEKILSA 1208

Db 1145 GYKIIIRKIRGNTNDRVRGGDILYFDMMINKAYNLFMKNETMYADNHSTEDIYALGLR 1204

Qy 1209 LEIPDVGNIQSQVVMKSKNDQGITNKC-KXNLODNNNGNDIGFTGFHFQ-----NNI 1258

Db 1205 EQTKDINDNIIFIQIPMNNTYYVASQIFKSNFNGENISGICSTGYRFLRGSDWYRHNLY 1264

Qy 1259 AKLVASWYRQIERSRSLGCSWEIPIVDD 1289

Db 1265 VPTVKGQNYASLLESTs-----THWGFVPVSE 1291

RESULT 15

S48110

neurotoxin type F - Clostridium botulinum (fragment)

C:Species: Clostridium botulinum

C>Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C:Accession: S48110

R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulin neurotoxin gene and specific ide

A:Reference number: S48103; MUID:94013372; PMID:8408542

A:Accession: S48110

A:Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 1-366 <CAM>

A:Cross-references: UNIPROT:Q57236; EMBL:X70821; NID:G407792; PIDN:CAA50152.1; PID:G40779

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 12.4%; Score 847; DB 2; Length 366;

Best Local Similarity 43.7%; Pred. No. 3.7e-30;

Matches 159; Conservative 85; Mismatches 114; Indels 6; Gaps 3;

Qy 649 KDDPVGALIFSGAVILLEPIETAIPVLGTFALVSYTA---NKVLTQITDIALSKRNEK 705

Db 3 KENFKEAFELGAGILLEFVPELLIPTILVFTIKSGSENKNKIKAINNLSMERETK 62

Qy	706	WDEVYKIVTNWIAKVNTOQDLIRKXKWEALENOABATKAIINYQNOYTEEEKNNI--N	763
Db	63	WKEIYSWIVSNWLTTRINTQFNKRKEQMYQALQONQVDAIKTVIEYKKNYTSDERNRLESE	122
Qy	764	FNIDDLSSKLNESINKAMININKFNLNOCVSYLMNSMIPYGVKRLDPDASLKDALLKYI	823
Db	123	YNINNIREELNKKVSLAMENIERPITESSIFYLMKLINEAKVSKLREYDEGVKEYLLDYI	182
Qy	824	YDNRGTLIQVDRLKDKVNNVTLSTDIPFOLSXYVDNQRLSLTFTYIKNIINTSIILNRY	883
Db	183	SEHRSILGNSVQELNDLVTSLNNSIPFELSSYTNDKILILYFNKLYKKIKDNSILDMRY	242
Qy	884	ESNHLIDLSRYASKINTGSKVNFDPIDKNOIQLFNLESSKIEVILKNAIYVNSMYENFST	943
Db	243	ENNKFDISGYGNSINGDVYIYSTYSTRNQFGIYSSKPSVNTAQNNDIYNGRYQNFSI	302
Qy	944	SPWIRIPKYENSISLANEYTIINCM-ENNSGWKVSILNYGEIITWTLODTQEIQRVVEKYS	1002
Db	303	SPWVRIPKPYFNKVLNNEYTIIDCIRNNSGWKISLNYNKIITWTLODTAGNNOKLVFNFT	362
Qy	1003	QMIN 1006	
Db	363	QMIS 366	

Search completed: January 31, 2005, 14:09:14
Job time : 65 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:03 ; Search time 224 Seconds
(without alignments)
3344.365 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFVNQFNKDPVNGVDIA.....EFIPVDDGWERPLHHHHH 1302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6768	99.2	1296	2	AA075961
2	6768	99.2	1296	2	AAQ06331
3	6759	99.1	1295	1	BXA1_CLOBO
4	6153	90.2	1295	1	BXA2_CLOBO
5	2504.5	36.7	1268	2	Q45851
6	2431	35.6	1278	2	Q57236
7	2368.5	34.7	1251	2	Q9K395
8	2339	34.3	1252	2	BAB86845
9	2332	34.2	1252	2	Q8KZM3
10	2332	34.2	1255	2	Q9FAR6
11	2319.5	34.0	1250	1	BXE_CLOBO
12	2310	33.9	448	2	Q84G74
13	2309.5	33.9	1250	1	BXE_CLOBO
14	2296.5	33.7	1274	1	BXF_CLOBO
15	2290	33.6	1291	2	Q8GR36
16	2283.5	33.5	1280	2	Q9ZAJ5
17	2279	33.4	1296	1	BXG_CLOBO
18	2267	33.2	1291	2	Q08077
19	2266	33.2	1290	1	BXB_CLOBO
20	2258	33.1	1291	2	Q9ZAJ8
21	2258	33.1	1291	2	Q93G71
22	2255	33.1	1291	2	Q933K0
23	1890.5	27.7	1276	1	BXD_CLOBO
24	1888.5	27.7	1275	2	Q9QTG7
25	1887.5	27.7	1314	1	TETX_CLOTE
26	1854.5	27.2	1310	2	Q93N27
27	1851	27.1	1285	2	Q45967
28	1846	27.1	1285	2	Q9LBR1
29	1781.5	26.1	1280	2	Q9LBS7
30	1777.5	26.1	1280	2	Q45849
31	1774.5	26.0	1280	2	Q84153

32	1771	26.0	1291	2	Q93HT3	Q93ht3 clostridium
33	1764	25.9	1290	1	BXC1_CLOBO	P18640 clostridium
34	1360	19.9	260	2	Q45892	Q45892 clostridium
35	921.5	13.5	361	2	Q45846	Q45846 clostridium
36	918.5	13.5	361	2	Q45848	Q45848 clostridium
37	847	12.4	366	2	Q79AH9	Q79ah9 clostridium
38	818.5	12.0	441	2	Q9X708	Q9x708 clostridium
39	808.5	11.9	367	2	Q45862	Q45862 clostridium
40	805.5	11.8	367	2	Q45861	Q45861 clostridium
41	651	9.5	1198	2	Q06018	Q06018 clostridium
42	636	9.3	1197	2	P71117	P71117 clostridium
43	634	9.3	1197	2	Q33871	Q33871 clostridium
44	633.5	9.3	451	2	Q9LA13	Q9la13 clostridium
45	628	9.2	1197	2	Q9ZAJ9	Q9zaJ9 clostridium

ALIGNMENTS

RESULT 1
AA075961
ID AA075961 PRELIMINARY; PRT; 1296 AA.
AC AA075961;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE BONT/A.
GN BONT/A.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hall A-byder;
RX MEDLINE=22617869; PubMed=12732962;
RA Dineen S.S., Bradshaw M., Johnson E.A.;
RT "Neurotoxin Gene Clusters in Clostridium botulinum Type A Strains:
Sequence Comparison and Evolutionary Implications.";
RL Curr. Microbiol. 46:345-352(2003).
DR EMBL: AF461540; AA075961.1; -
SQ SEQUENCE 1296 AA; 149425 MW; DEAC2F2754AE43E6 CRC64;

Query Match	99.2%	Score	6768	DB	2	Length	1296
Best Local Similarity	99.9%	Pred. No.	2.2e-293				
Matches	1295	Conservative	1	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MPFVNQFNKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIIPERDTFTNPEEGDLN	60				
Db	1	MPFVNQFNKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIIPERDTFTNPEEGDLN	60				
Qy	61	PPPEAKQVPVSYDSTYLSVDNEKDYKLVGVTGLFERIYSTDLGRMLLTISIVRGIPFWGG	120				
Db	61	PPPEAKQVPVSYDSTYLSVDNEKDYKLVGVTGLFERIYSTDLGRMLLTISIVRGIPFWGG	120				
Qy	121	STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY	180				
Db	121	STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY	180				
Qy	181	GSTQYIRFSPDPFTGFEESLEVDTNPLLGAGKATDPAVTLAHLIYAGHRLYGIATNP	240				
Db	181	GSTQYIRFSPDPFTGFEESLEVDTNPLLGAGKATDPAVTLAHLIYAGHRLYGIATNP	240				
Qy	241	RVFKVNTNAYEMSGLEVSPFEELRTFGCHDAKFTDSIQENEFRLYYNKKFKDIASLTNKA	300				
Db	241	RVFKVNTNAYEMSGLEVSPFEELRTFGCHDAKFTDSIQENEFRLYYNKKFKDIASLTNKA	300				
Qy	301	KSIVGTTASIQYMKNVKPKYLLSEDTSGKFSVDKLFKDKLXKMLTEIYTEDNFVKFKV	360				
Db	301	KSIVGTTASIQYMKNVKPKYLLSEDTSGKFSVDKLFKDKLXKMLTEIYTEDNFVKFKV	360				
Qy	361	LNKRTYLNFKPAKFNIVPKVNYTYDGNLNTNLAAFPNGQNTENNWNFKLNFT	420				
Db	361	LNKRTYLNFKPAKFNIVPKVNYTYDGNLNTNLAAFPNGQNTENNWNFKLNFT	420				

Db 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480
Db 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480
Qy 481 ITSDTNIEAABENISLDLIQYYITFNFDPNENISNIENSSDIIGOLELMPNIEFPNG 540
Db 481 ITSDTNIEAABENISLDLIQYYITFNFDPNENISNIENSSDIIGOLELMPNIEFPNG 540
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Db 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPAINIGNMLYKDDFVGCALIFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPAINIGNMLYKDDFVGCALIFSG 660
Qy 661 AVILLEPIEIPAIPVLGTFALVSIVANKVLTQVOTIDNALS KRNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLEPIEIPAIPVLGTFALVSIVANKVLTQVOTIDNALS KRNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAATKAIINYOQNYTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAATKAIINYOQNYTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCSSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGVDRLKDK 840
Db 781 MININKFLNOCSSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGVDRLKDK 840
Qy 841 VNNTLSTDIPFQLSKYVDNORLLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI 900
Db 841 VNNTLSTDIPFQLSKYVDNORLLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI 900
Qy 901 GSKVNFDPIDKXQIQLFNLESKEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKXQIQLFNLESKEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTIINCMEANS GWKSVLNYGEIITWLTQDTEIKQVVFVKYSQMINISDIYNRWIFVTIT 1020
Db 961 EYTIINCMEANS GWKSVLNYGEIITWLTQDTEIKQVVFVKYSQMINISDIYNRWIFVTIT 1020
Qy 1021 NNRNNSKIYNGRLIDQKPTISNLGNHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRNNSKIYNGRLIDQKPTISNLGNHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKDLYDNOSGILKDFWGDYLOYDKPYMLNLVDPNKYVDVNNVGIRGYMLKQPR 1140
Db 1081 EKEIKDLYDNOSGILKDFWGDYLOYDKPYMLNLVDPNKYVDVNNVGIRGYMLKQPR 1140
Qy 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
Db 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALETPDVGNLSQVVMKSKNDQGTNCKMNLQDNNGNDIGCFGHFQFNNAK 1260
Db 1201 GVEKILSALETPDVGNLSQVVMKSKNDQGTNCKMNLQDNNGNDIGCFGHFQFNNAK 1260
Qy 1261 LVASNNYRQIERSRTLGCSWEFIPVDDGGERPL 1296
Db 1261 LVASNNYRQIERSRTLGCSWEFIPVDDGGERPL 1296

RESULT 2

AAQ06331 PRELIMINARY; PRT; 1296 AA.
AC AAQ06331;
DT 02-MAR-2004 (TremBLrel. 27, Created)
DT 02-MAR-2004 (TremBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TremBLrel. 27, Last annotation update)
DE Neurotoxin BoNT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Allergan-Hall A;
RX MEDLINE=22919384; PubMed=14557061;
RA Zhang L., Lin W.J., Li S., Aoki K.R.;
RT "Complete DNA sequences of the botulinum neurotoxin complex of
Clostridium botulinum type A-Hall (Allergan) strain.";
RL Gene 315:21-32(2003).
DR EMBL; AP488749; AAQ06331.1; -.
KW Neurotoxin.
SQ SEQUENCE 1296 AA; 149425 MW; DEACF2754AE43E6 CRC64;

Query Match 99.2%; Score 6768; DB 2; Length 1296;

Best Local Similarity 99.9%; Pred. No. 2.2e-293;

Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPFVNKQFNKYKDPVNGVDIAYIKIPNAGQMPKAFKIHNNKIWIPIERDFTFTNPEGDNL 60
Db 1 MPFVNKQFNKYKDPVNGVDIAYIKIPNAGQMPKAFKIHNNKIWIPIERDFTFTNPEGDNL 60
Qy 61 PPPEAKQVPVSYVDSTVSLTDNEKNYKGVTKLFEIYSTDLGRMLLTISIVRGIPWGG 120
Db 61 PPPEAKQVPVSYVDSTVSLTDNEKNYKGVTKLFEIYSTDLGRMLLTISIVRGIPWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGVSRSSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGVSRSSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKATDPAVTLAHLIYVAGHRLYGIAINPN 240
Db 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKATDPAVTLAHLIYVAGHRLYGIAINPN 240
Qy 241 RVKVNNTNAYVSGLEVSEELRTFGCHDAKFTDSLOENEFRLYYNKKFKDASTLANKA 300
Db 241 RVKVNNTNAYVSGLEVSEELRTFGCHDAKFTDSLOENEFRLYYNKKFKDASTLANKA 300
Qy 301 KSIIVGTTASIQYMKNVFKEKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYTDNPFVKFKV 360
Db 301 KSIIVGTTASIQYMKNVFKEKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYTDNPFVKFKV 360
Qy 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Db 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480
Db 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480
Qy 481 ITSDTNIEAABENISLDLIQYYITFNFDPNENISNIENSSDIIGOLELMPNIEFPNG 540
Db 481 ITSDTNIEAABENISLDLIQYYITFNFDPNENISNIENSSDIIGOLELMPNIEFPNG 540
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Db 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPAINIGNMLYKDDFVGCALIFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPAINIGNMLYKDDFVGCALIFSG 660
Qy 661 AVILLEPIEIPAIPVLGTFALVSIVANKVLTQVOTIDNALS KRNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLEPIEIPAIPVLGTFALVSIVANKVLTQVOTIDNALS KRNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAATKAIINYOQNYTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAATKAIINYOQNYTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCSSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGVDRLKDK 840
Db 781 MININKFLNOCSSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLLIGVDRLKDK 840

Qy 841 VVNTLSTDDIPOLSKYVDNQRLSTFTTEYIKNIINTSILMLRYESNHLIDLSRYASKINI 900
 Db 841 VVNTLSTDDIPOLSKYVDNQRLSTFTTEYIKNIINTSILMLRYESNHLIDLSRYASKINI 900
 Qy 901 GSKVNFDPIDKNQIQLEFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
 Db 901 GSKVNFDPIDKNQIQLEFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
 Qy 961 EYTIINCENNSGKWSLNTGEEIIWTLDQTOEIKQIRVVFYKYSQMINISDYINRWIFVTIT 1020
 Db 961 EYTIINCENNSGKWSLNTGEEIIWTLDQTOEIKQIRVVFYKYSQMINISDYINRWIFVTIT 1020
 Qy 1021 NNRLNNSKIYINGRLIDQKPSINLGNTHASNNIMFKLDGGRDTHRYIWIYFNLFDEKELN 1080
 Db 1021 NNRLNNSKIYINGRLIDQKPSINLGNTHASNNIMFKLDGGRDTHRYIWIYFNLFDEKELN 1080
 Qy 1081 EKETIKDYDQNSGIIKDFWGDYLOVDKPYVLMNLVDPNKKYVDVNVNGVIRGYMYLKGPR 1140
 Db 1081 EKETIKDYDQNSGIIKDFWGDYLOVDKPYVLMNLVDPNKKYVDVNVNGVIRGYMYLKGPR 1140
 Qy 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
 Db 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
 Qy 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKVNLQDNNGNDIGFIFGHQFNNAK 1260
 Db 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKVNLQDNNGNDIGFIFGHQFNNAK 1260
 Qy 1261 LVASNNYNRQIERSRSLTGCSWEFIPVDGNGERPL 1296
 Db 1261 LVASNNYNRQIERSRSLTGCSWEFIPVDGNGERPL 1296

RESULT 3

EXAL_CLOBO STANDARD; PRT; 1295 AA.
 AC P10845; P01561; P18639;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-chain; Botulinum neurotoxin A, heavy-chain].
 GN Name=botA; Synonym=bna, atx;
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Type A / NCTC 2916;
 RX MEDLINE=90235864; PubMed=2185020;
 RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J., Shone C.C.,
 RA Atkinson T., Melling J., Minton N.P.;
 RT "The complete amino acid sequence of the Clostridium botulinum type A
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
 RT gene.";
 RL Eur. J. Biochem. 189:73-81(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Type A / 62A;
 RX MEDLINE=90264400; PubMed=2160960;
 RA Binz B., Kuarzono H., Willie M., Frevent J., Wernars K., Niemann H.;
 RT "The complete sequence of botulinum neurotoxin type A and comparison
 RT with other clostridial neurotoxins.";
 RL J. Biol. Chem. 265:9153-9158(1990).
 RN [3]
 RN SEQUENCE OF 1-65 FROM N.A.
 RC STRAIN=Type A / 62A;
 RX MEDLINE=97016817; PubMed=8863443;
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
 RT "Organization and phylogenetic interrelationships of genes encoding

RT components of the botulinum toxin complex in proteolytic Clostridium
 RT botulinum types A, B, and F: evidence of chimeric sequences in the
 RT gene encoding the nontoxic nonhemagglutinin component.";
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
 RN [4]
 RN SEQUENCE OF 1-34 FROM N.A.
 RC STRAIN=Type A / Hall;
 RX MEDLINE=89350959; PubMed=2669749;
 RA Betley M.J., Somers E., Dasgupta B.R.;
 RT "Characterization of botulinum type A neurotoxin gene: delineation of
 RT the N-terminal encoding region.";
 RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
 RN [5]
 RN SEQUENCE OF 1-18 FROM N.A.
 RC STRAIN=Type A / NIH;
 RX MEDLINE=96096783; PubMed=8521962;
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
 RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin
 RT components of Clostridium botulinum type A progenitor toxins.";
 RL FEBS Lett. 376:41-44(1995).
 RN [6]
 RN SEQUENCE OF 1-16.
 RP MEDLINE=84178501; PubMed=6370252;
 RX Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
 RA "Partial amino acid sequence of the heavy and light chains of
 RT botulinum neurotoxin type A.";
 RL Biochem. Biophys. Res. Commun. 119:900-904(1984).
 RN [7]
 RN SEQUENCE OF 1-46.
 RP Dasgupta B.R., Foley J., Niece R.;
 RA "Partial sequence of the light chain of botulinum neurotoxin type A.";
 RL Biochemistry 26:4162-4162(1987).
 RN [8]
 RN SEQUENCE OF 1-5 AND 444-456.
 RX MEDLINE=91120847; PubMed=2126206;
 RA Dasgupta B.R., Dekleva M.L.;
 RT "Botulinum neurotoxin type A: sequence of amino acids at the N-
 RT terminus and around the nicking site.";
 RL Biochimie 72:661-664(1990).
 RN [9]
 RN SEQUENCE OF 448-464 AND 872-895.
 RX MEDLINE=89024662; PubMed=3178218;
 RA Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
 RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two
 RT halves and their partial sequences.";
 RL Arch. Biochem. Biophys. 266:142-151(1988).
 RN [10]
 RN SEQUENCE OF 448-482.
 RP MEDLINE=85285016; PubMed=3896784;
 RX Shone C.C., Hambleton P., Melling J.;
 RA "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
 RT and purification of two tryptic fragments. Proteolytic action near the
 RT COOH-terminus of the heavy subunit destroys toxin-binding activity.";
 RL Eur. J. Biochem. 151:75-82(1985).
 RN [11]
 RN SEQUENCE OF 866-879 AND 1147-1218.
 RP Gimenez J.A., Dasgupta B.R.;
 RX "Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72,
 RT 45, 42, and 18 kD fragments.";
 RL J. Protein Chem. 12:351-363(1993).
 RN [12]
 RN IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94063091; PubMed=8243676;
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
 RA Benfenati F., Wilson M.C., Montecucco C.;
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
 RT COOH-terminal peptide bonds.";
 RL FEBS Lett. 335:99-103(1993).
 RN [13]
 RN IDENTIFICATION OF SUBSTRATE.
 RP MEDLINE=94124495; PubMed=8294407;
 RX Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,

RA Jahn R., Niemann H.;
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
 RL J. Biol. Chem. 269:1617-1620(1994).
 RN [14]
 RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
 RX MEDLINE=2155894; PubMed=11700044; DOI=10.1006/bbrc.2001.5911;
 RA Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
 RT "Site-directed mutagenesis identifies active-site residues of the
 RL light chain of botulinum neurotoxin type a.";
 RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
 RN [15]
 RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
 RX MEDLINE=94455071; PubMed=9783750;
 RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
 RT "Crystal structure of botulinum neurotoxin type A and implications for
 RL toxicity";
 RL Nat. Struct. Biol. 5:898-902(1998).
 CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
 CC binds with high affinity to peripheral neuronal presynaptic
 CC membrane, is then internalized by receptor-mediated endocytosis.
 CC The C-terminus of the heavy chain (H) is responsible for the
 CC adherence of the toxin to the cell surface while the N-terminus
 CC mediates transport of the light chain from the endocytic vesicle
 CC to the cytosol. After translocation, the light chain (L)
 CC hydrolyzes the 197-Gln-Arg-198 bond in SNAP-25, thereby blocking
 CC neurotransmitter release. Inhibition of acetylcholine release
 CC results in flaccid paralysis, with frequent heart or respiratory
 CC failure.
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
 CC detected action on small molecule substrates.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit.
 CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
 CC heavy chain (H).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for the
 CC treatment of strabismus and blepharospasm associated with dystonia
 CC and cervical dystonia. Also used for the treatment of hemifacial
 CC spasm and a number of other neurological disorders characterized
 CC by abnormal muscle contraction.
 CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of
 CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
 CC -1- SIMILARITY: Belongs to peptidase family M27.
 CC -1- DATABASE: NAME=BOTOX product information Web site;
 CC WWW="http://www.botox.com/site/";
 CC -1- DATABASE: NAME=Protein Spotlight; NOTE=Issue 19 of February 2002;
 CC WWW="http://www.expasy.org/spotlight/articles/spl1019.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52066; CAA36289.1; -
 CC EMBL; M30196; AAA23262.1; -
 CC EMBL; X92973; CAA63551.1; -
 CC EMBL; D67030; BAA11051.1; -
 CC EMBL; M27892; AAA23269.1; -
 CC PIR; A35294; BTCLAB.
 CC PDB; 3BTA; X-ray; A=1-1295.
 CC MEROPS; M27.002; -
 CC InterPro; IPR008985; ConA like lec_gl.
 CC InterPro; IPR011065; Kunitz like.
 CC InterPro; IPR003395; Peptidase M27.
 CC InterPro; IPR006025; Pept M Zn BS.
 CC Pfam; PF01742; Peptidase M27; 1.
 CC PRINTS; PR00760; BONTOXILYSIN.
 CC ProDom; PD001963; Bontoxilysin; 1.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC 3D-structure; Direct protein sequencing; Hydrolase; Metalloprotease;
 KW

KW Neurotoxin; Pharmaceutical; Transmembrane; Zinc.
 FT INIT MET 0
 FT CHAIN 1 447 Botulinum neurotoxin A light-chain.
 FT CHAIN 448 1295 Botulinum neurotoxin A heavy-chain.
 FT METAL 222 222 Zinc (catalytic).
 FT ACT_SITE 223 223 Zinc (catalytic).
 FT METAL 226 226 Zinc (catalytic).
 FT METAL 261 261 Zinc (catalytic).
 FT DISULFID 429 453 Interchain.
 FT TRANSMEM 1234 1279 Potential.
 FT TRANSMEM 626 646 Potential.
 FT VARIANT 26 26 V -> A.
 FT MUTAGEN 261 261 E->A: Drastic decrease in enzymatic activity.
 FT MUTAGEN 265 265 F->A: Decreases enzymatic activity.
 FT MUTAGEN 365 365 Y->A: Decreases enzymatic activity.

Query Match 99.1%; Score 6759; DB 1; Length 1295;
 Best Local Similarity 99.8%; Pred. No. 5.5e-293;
 Matches 1293; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PFVKNQNYKDPVNGVDIAVIKIPNAGQMPQVAKFIHNKIWIPIPERDTFTNPEEGDLP 61
 DB 1 PFVKNQNYKDPVNGVDIAVIKIPNAGQMPQVAKFIHNKIWIPIPERDTFTNPEEGDLP 60
 QY 62 PPEAKQVPVSYDYSTLSTDNKDNLYKGVTKLFIERYSTDLGRMLLTSTIVRGIPFWGGS 121
 DB 61 PPEAKQVPVSYDYSTLSTDNKDNLYKGVTKLFIERYSTDLGRMLLTSTIVRGIPFWGGS 120
 QY 122 TIDTELKVIDTNCINVIQDPGYSRSEELNVIIGPSADIIQFCKSFHGVNLTRNGY 181
 DB 121 TIDTELKVIDTNCINVIQDPGYSRSEELNVIIGPSADIIQFCKSFHGVNLTRNGY 180
 QY 182 STQVIRSPDFTGRFEESLEVDNPLLGAGKFAIDPAVTLAHLIYAGHLYGAIINPNR 241
 DB 181 STQVIRSPDFTGRFEESLEVDNPLLGAGKFAIDPAVTLAHLIYAGHLYGAIINPNR 240
 QY 242 VFQVNTNAYYEMSGLEVSPFEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKAK 301
 DB 241 VFQVNTNAYYEMSGLEVSPFEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKAK 300
 QY 302 SIQVTTASLQYMKVFKKYLSEDSGKFSVDKLFKDKLYKMLTEIYTDNFKVFFKVL 361
 DB 301 SIQVTTASLQYMKVFKKYLSEDSGKFSVDKLFKDKLYKMLTEIYTDNFKVFFKVL 360
 QY 362 NRKTYLNFDKAVKINIVPKVNTIYDGFNLRTNLAAFNQNTNINNNFTKLNFTG 421
 DB 361 NRKTYLNFDKAVKINIVPKVNTIYDGFNLRTNLAAFNQNTNINNNFTKLNFTG 420
 QY 422 LPEFYKLLCVRGITTSKTSKLDKGYNKALNDLCIKVNNWDLFPSPSDNFTNDLNKGEI 481
 DB 421 LPEFYKLLCVRGITTSKTSKLDKGYNKALNDLCIKVNNWDLFPSPSDNFTNDLNKGEI 480
 QY 482 TSDNTIEAARENISLDLIQOYLYTFNFDNEPENISLENISDIIQGLLELMPNTERFNGK 541
 DB 481 TSDNTIEAARENISLDLIQOYLYTFNFDNEPENISLENISDIIQGLLELMPNTERFNGK 540
 QY 542 KYELDQVTFMTHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 601
 DB 541 KYELDQVTFMTHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 600
 QY 602 MFLGWVQLVYDFTDETSVSTTKDIATITIIPIYGPALNIGMLYKDDFVGALIFSGA 661
 DB 601 MFLGWVQLVYDFTDETSVSTTKDIATITIIPIYGPALNIGMLYKDDFVGALIFSGA 660
 QY 662 VILLEFPEITAI PVLGTFALVSYANKVLVOTIDNALSKEKNEKDWVYKIVTNWLAKV 721
 DB 661 VILLEFPEITAI PVLGTFALVSYANKVLVOTIDNALSKEKNEKDWVYKIVTNWLAKV 720
 QY 722 NTQIDLIRKQKALENOAEATKAIINYQVNTYEEKNNINFNIDDLSSKLNESINKAM 781
 DB 721 NTQIDLIRKQKALENOAEATKAIINYQVNTYEEKNNINFNIDDLSSKLNESINKAM 780

QY 782 ININKELNQC SVSYLMSMTPYGVRLLEDPSALKDALLKYIYVDRGTLIGQVDRKDKV 841
DB 781 ININKELNQC SVSYLMSMTPYGVRLLEDPSALKDALLKYIYVDRGTLIGQVDRKDKV 840
QY 842 NNTLSTDI PPQASKYVDNQLLSTFEYIKNIINTSILNRYESNHLIDLRSYASKINIG 901
DB 841 NNTLSTDI PPQASKYVDNQLLSTFEYIKNIINTSILNRYESNHLIDLRSYASKINIG 900
QY 902 SKYFDPIDKQQLFNLESSKIEVLKNAIVNYSNMFSTFWIRIPKYNFISILNNE 961
DB 901 SKYFDPIDKQQLFNLESSKIEVLKNAIVNYSNMFSTFWIRIPKYNFISILNNE 960
QY 962 YTIINCHENNKGWKSILNYGELIITWLODTEIKQVVFVKYSOMINSDYINRWIFVTITN 1021
DB 961 YTIINCHENNKGWKSILNYGELIITWLODTEIKQVVFVKYSOMINSDYINRWIFVTITN 1020
QY 1022 NRLNNSKIYINGRLIDOKPISNIGNTHASNNIMFKDGCDDTHRYIWKYFNLFDKELNE 1081
DB 1021 NRLNNSKIYINGRLIDOKPISNIGNTHASNNIMFKDGCDDTHRYIWKYFNLFDKELNE 1080
QY 1082 KEIKDLYDNQSGILKDFWGDYLDYDKPYMLNLYDPNKYVDVNVNNGIRGYMYLKGPRG 1141
DB 1081 KEIKDLYDNQSGILKDFWGDYLDYDKPYMLNLYDPNKYVDVNVNNGIRGYMYLKGPRG 1140
QY 1142 SVMTTNIYLNSSLYRGTKPIIKKYASGNKDNIVRNDRVYINVVVKYRVLATNASQAG 1201
DB 1141 SVMTTNIYLNSSLYRGTKPIIKKYASGNKDNIVRNDRVYINVVVKYRVLATNASQAG 1200
QY 1202 VEKILSALETPDVGNSLQVVMKSKNDQGTNCKMNLQNDNGDIFGIFGHFNNTAKL 1261
DB 1201 VEKILSALETPDVGNSLQVVMKSKNDQGTNCKMNLQNDNGDIFGIFGHFNNTAKL 1260
QY 1262 VASWYNRQIERSRSLGCSWEFIPVDDGMGERPL 1296
DB 1261 VASWYNRQIERSRSLGCSWEFIPVDDGMGERPL 1295

RESULT 4
BXA2_CLOBO STANDARD; PRT; 1295 AA.
AC Q45894; P77780;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-chain; Botulinum neurotoxin A, heavy-chain].
DE Name=botA; Synonyms=bna, atx;
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Type A / Kyoto-F;
RX MEDLINE=94143603; PubMed=8310180;
RA Wallens A., East A.K., Lawson P.A., Collins M.D.;
RT "Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins.";
RL Res. Microbiol. 144:547-556(1993).
RN [2]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=Type A / Kyoto-F;
RX MEDLINE=37016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic Clostridium botulinum types A, B, and F: evidence of chimeric sequences in the RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin

CC binds with high affinity to peripheral neuronal presynaptic membrane, is then internalized by receptor-mediated endocytosis. The C-terminus of the heavy chain (H) is responsible for the adherence of the toxin to the cell surface while the N-terminus mediates transport of the light chain from the endocytic vesicle to the cytosol. After translocation, the light chain (L) hydrolyzes the 197-Gln-Arg-198 bond in SNAP-25, thereby blocking neurotransmitter release. Inhibition of acetylcholine release results in flaccid paralysis, with frequent heart or respiratory failure (By similarity).
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.
CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a heavy chain (H) (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -1- SIMILARITY: Belongs to peptidase family M27.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X73423; CAA51824.1; -;
DR EMBL; X87974; CAA61234.1; -;
DR PIR; I40645; I40645.
DR PDB; 1E1H; X-ray; A/Cs1-249, B/D=250-423.
DR MEROPS; M27.002; -;
DR InterPro; IPR008985; ConA like lec_g1.
DR InterPro; IPR011065; Kunitz like.
DR InterPro; IPR000395; Peptidase M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC PROTEASE; FALSE NEG.
DR 3D-structure; Hydrolase; Metalloprotease; Neurotoxin; Transmembrane; Zinc.
KW INIT_MET 0 0 By similarity.
KW CHAIN 1 447 Botulinum neurotoxin A light-chain.
KW CHAIN 448 1295 Botulinum neurotoxin A heavy-chain.
FT METAL 222 222 Zinc (Catalytic) (By similarity).
FT ACT_SITE 223 223 By similarity.
FT METAL 226 226 Zinc (catalytic) (By similarity).
FT DISULFID 429 453 Interchain (By similarity).
FT TRANSMEM 626 646 Potential.
FT TRANSMEM 655 675 Potential.
SQ SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;

Query Match 90.2%; Score 6153; DB 1; Length 1295;
Best Local Similarity 89.9%; Pred. No. 5.7e-266;
Matches 1164; Conservative 66; Mismatches 65; Indels 0; Gaps 0;

QY 2 PFVKNQFNYKDPVNGVDIAYIKIPNAGOMQPVAFKHNKIWIIPERDTFTNPEEGDLP 61
DB 1 PFVKNQFNYKDPVNGVDIAYIKIPNAGOMQPVAFKHNKIWIIPERDTFTNPEEGDLP 60

QY 62 PPEAKQVPVSYDYSTYLTSTNEDKNYLGKVTLPRIYSTDGLGMLLTSTVIRGIPFWGGS 121
DB 61 PPEAKQVPVSYDYSTYLTSTNEDKNYLGKVTLPRIYSTDGLGMLLTSTVIRGIPFWGGS 120

QY 122 TIDTELKVIDTNCINVIQPDGYSRSEELNVLVIIGPSADIIQFECKSFGEVNLNTRNGY 181
DB 121 TIDTELKVIDTNCINVIQPDGYSRSEELNVLVIIGPSADIIQFECKSFGEVNLNTRNGY 180

QY 182 STQYIRSPDFTGFBRESLEVDNPNLLGAGKPAIDPAVTLAHELIYAGHLYGTAIPNR 241
DB 181 STQYIRSPDFTGFBRESLEVDNPNLLGAGKPAIDPAVTLAHELIYAGHLYGTAIPNR 240

```
Db 181 STQVIRFSPDFTFGFEESLEVDNTPLLGAGKFAATDPATVTLAHELIIHAHRLYGTAIINPNR 240
QY 242 VFKNNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENFRLYYNKFQDIASLTAKAK 301
Db 241 VFKNNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENFRLYYNKFQDVASTLAKAK 300
QY 302 SIVGTTASLOVMKNVFKKYLSDTSKGSVDKLFKFDKLYKMLTEIYTDNFVFKFKVL 361
Db 301 SIIGTTASLOVMKNVFKKYLSDTSKGSVDKLFKFDKLYKMLTEIYTDNFVFNFKVI 360
QY 362 NRKTYLNFDAKAVKINIVPKVNYIYDGFNLRTNLAANFNGQTEINNWFYKLNKFTG 421
Db 361 NRKTYLNFDAKAVFRINIVPDENYITKGFNLKGANLSTFNGQTEINSRFTLKNFTG 420
QY 422 LFEPYKLLCVRGITTSKTKSLDKGNKALNDLCIKVNNWDLFPSPSDNFTNDLNGKEEI 481
Db 421 LFEPYKLLCVRGIIIPFTKSLDEGNKALNDLCIKVNNWDLFPSPSDNFTNDLKVVEI 480
QY 482 TSDTNIBAAENISLDLIQOYYLTFNFDNBPENISIEIENLSSDIIGOLELMPNIEFPNGK 541
Db 481 TADTNIBAAENISLDLIQOYYLTFNFDNBPENISIEIENLSSDIIGOLELMPNIEFPNGK 540
QY 542 KYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 601
Db 541 KYELDKYTMFHYLRAQFEHGKSRILITNSAAEALLKPNVAYTFSSKYVKKNKAVEAF 600
QY 602 MFLGWVQLVYDFTDETSEVSTDKIADITIIPIYIGPALNIGMLYKDDFVGALIFSGA 661
Db 601 MFLNWAELVYDFTDETNEVTMOKIADITIIPIYIGPALNIGMLSKGFEVEAIIITGV 660
QY 662 VILLEFIPEIAIPVLGTFALVSYTANKVLTVQTTIDNALSKEKDEWYKYIYVTNWLAKV 721
Db 661 VAMLEFIPEYALPVGTFALVSYTANKVLTVQTTINNALSKEKDEWYKYIYVTNWLAKV 720
QY 722 NTQIDLRKMKALENOAEATKAIINYQNYTEEEKNNINFNIDLSKLINESINKAM 781
Db 721 NTQIDLRKMKALENOAEATKAIINYQNYTEEEKNNINFNIDLSKLINESINSAM 780
QY 782 ININKFLNQCSVSLNMSMIPYGVKRLDFDASLKDALLKYIYDNRGTIGQVDRLKDKV 841
Db 781 ININKFLDQCSVSLNMSMIPYAVKRLKDFDASVDRVLLKYIYDNRGTILVQVDRLKDEV 840
QY 842 NNTLSTDIPQLSKYVDNORLLSTFTYIKNIINTSILNRLYESNHLIDLSRYASKINIG 901
Db 841 NNTLSADIPQLSKYVDNKKLLSTFTYIKNIINTSILSVYKDDLDLSRYGAKINIG 900
QY 902 SKYNFDPIDKNOIQLFNLESSKIEVILKNAIVNMSYENFSTSPWIRIPKYFNSISLANE 961
Db 901 DRVYDSIDKXNQIKLNLSTIEVILKNAIVNMSYENFSTSPWIKIPKYFSKINLANE 960
QY 962 YTIINCMEENSGKVSINLYGELIITLQDTQBIKQVVFVKYSQMINISDIYNRMIFVITN 1021
Db 961 YTIINCMEENSGKVSINLYGELIITLQDNQNIQVVFVKYSQMINISDIYNRMIFVITN 1020
QY 1022 NRLNNSKIYINGRLIDQKIPISNLGIIHASNNIMFKLDGCRDTHRYIYIKVFNLPDKELNE 1081
Db 1021 NRTLTKSIYINGRLIDQKIPISNLGIIHASNNIMFKLDGCRDPRYIYIKVFNLPDKELNE 1080
QY 1082 KEIKOLDYDQNSGILKDPFGDYLOYDKPYMLNLYDPNKYVDVNNVGIKGYMLKGPGRG 1141
Db 1081 KEIKOLDYDQNSGILKDPFGNYLOYDKPYMLNLYDPNKYVDVNNVGIKGYMLKGPGRG 1140
QY 1142 SVMNTNLYNLSLYRGTFIKKYASGNKONIVRNDRVYINVVVKNKEVRLATNASQAG 1201
Db 1141 SVVTNLYNLSLYRGTFIKKYASGNEDIVRNNDRVYINVVVKNKEVRLATNASQAG 1200
QY 1202 VEKILSALEIPDVGNLSQVVMVMSKNDQGITNKKCMKLDQNNGNDIGIFGHQNNIAKL 1261
Db 1201 VEKILSALEIPDVGNLSQVVMVMSKNDQGITNKKCMKLDQNNGNDIGIFGHLYDNIACL 1260
QY 1262 VASNNYRQIERSRRTIGCSWEFIPVDGGERPL 1296
Db 1261 VASNNYRQVGKASRTFGCSWEFIPVDGGESSL 1295
```

RESULT 5

```
Q45851
ID Q45851 PRELIMINARY; PRT; 1268 AA.
AC Q45851;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Neurotoxin type F.
GN Name:bont /i;
OS Clostridium baratii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93252228; PubMed=8486245;
RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium baratii type F
RT neurotoxin: comparison with other clostridial neurotoxins.";
RL FEMS Microbiol. Lett. 108:175-182(1993).
DR EMBL; X68262; CAA48329.1; -.
DR PIR; S33411; S33411.
DR HSSP; Q45894; IE1H.
DR MEROPS; M27.002; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0009405; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPRO11591; Botulinum.
DR InterPro; IPRO11065; Kunitz like.
DR InterPro; IPRO00395; Peptidase_M27.
DR InterPro; IPRO06025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BOTOXILYSIN.
DR ProDom; PD001963; Botulinum; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1268 AA; 145512 MW; 963040091AC15ED2 CRC64;
```

Query Match 36.7%; Score 2504.5; DB 2; Length 1268;
Best Local Similarity 41.4%; Pred. No. 2.4e-103;
Matches 552; Conservative 242; Mismatches 433; Indels 107; Gaps 36;

```
QY 1 MPFVKQFNKQDPVNGVDIAYIKIP-NAGQMPVKAFKHKHWIPIERTFTNPBGDL 59
Db 1 MPVNINNFNDPINNTIYMKPYYEDSNKYKAFKHMNDWIIIPERN-IIGKFSDF 59
QY 60 NPPEAKQVPSVYDSTYSLTDNEKNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWG 119
Db 60 YPISLDSGSSAYDPNYLTDAEKORFLKTVIKLFNRINSNPAGQVLLBEIKNGKPYLG 119
QY 120 GSTIDTELKVIDTNCIN-----VIQPDGYSSEELNVIIGPSADIIOPECKSPGHEV 172
Db 120 -----NDHTAVNEFCANNRSTSVIEKESNGCTTDSMLNLVLGPGPNIL--ECSTFFVRI 172
QY 173 L-----NLTRNGVGSYQYIRFSPDFTFGFEESLEVDNTPLLGAGKFAATPAVTLAHELI 226
Db 173 FPNIAVDPESEKFGSILQMSFSTFEYAFNDNTDL-----FIADPAISLAHELI 222
QY 227 YAGHRLYGI-AINPNRVKNTNAYVEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLY 285
Db 223 HVLHGLYGAGVTKKVKVIEVDQGLAM-AAEKDIKEBFITFGGQDLNIITNSTNQKIYVI 281
QY 286 YNKFQKDIASLT---NKAKSIVGTTASLQFMKNVFKKYLSDTSKGSVDKLFKPKLY 342
Db 282 LLSNYYTAIASRLSQVNRNNSALNTT-----YKKNFFQWKYGLDQDSNGNYTVNISKFNAIY 337
QY 343 KMLTEIYTDNFVKFKVLRKTYLNFDAKAVKINIVPKVNYIYDGFNLRTNLAANFN 402
Db 338 KKLIFS-FTECDLAQKQVKNRSNYLFFHKPFRLLDLDLDNIYISSEGFNI--GSLRVNNN 394
```


Db 434 LFFVASENSYNDNINTPKEIDDTVTNNYE-----NDLQVILNFSEAP-GLS 484
 Qy 517 IENLSSDIIGOLELMPNIERPENG-----KKYELDKYTMFHYLRAQEFHGKSRALTNSV 572
 Db 485 DEKLNLTQND-AVLPKYD--SNGTSDIEQHDVNLNVPFFYLDQAKVPEGENNVNLTSSI 541
 Qy 573 NEALLNPSRVYTFSSDYVKKVKNKATEAAMFLGWVQLVDFDTDETSEVSTDKIADITI 632
 Db 542 DTALLEQPKIYTFSSSEFINNVKPVQAALFVSWIQVVLDFTEANQKSTVDKIADISI 601
 Qy 633 IIPYIGPALNIGNMLYKDDPFGALIFSGAVILLFPIEIPAIPLVGTALVSYIA---NKV 689
 Db 602 WVPYIGLALNIGNEAQKGNFKDALELGAGILLFVPELLIPTILVTFIKSFLGSSDNKN 661
 Qy 690 LTVOTIDNALS KRNEKWDVYKYIVTNWLAKVNTQIDLRKMKKEALENOAEATKALINY 749
 Db 662 KVIRAINNALKERDEKKEVYFVSNWMTKINTQFNKRKEQMYQALQVNAKLTIEF 721
 Qy 750 QYNQYTEEEKNI--NFNIDDLSSKLNESINKAMINIKFNQCSVSYLNMSPYGVYK 807
 Db 722 KNSYTTLEEKELKNNYDIEQIENELNQKVSIAAMNIDRFLTESSISYLMKLINEVKINK 781
 Qy 808 LEDFDASLKDALLKYIYDNRGLTIGQVDR-LKDKVNNTLSTDIPFOLSKYVDNQRLISTF 866
 Db 782 LREYDENVKYLLNYIIQH-GSILGESQBELNSVTDLTNNISIPFKLSSYTDKILISYF 840
 Qy 867 TEYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEV 926
 Db 841 NKFFKRIKSSVLNMYKNDKYVDTSGYDSNINNGEIFYPTNKQFTFNSKPEVNI 900
 Qy 927 ILKNAIVNSWYENFSTFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSILNYGEII 984
 Db 901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNIINNEYTIINCMRDNNSGWKVSILNHEII 960
 Qy 985 WTLQDQOEIKQVVFVKYSOMINISDYINRWLFVTITNRLNNSKIYINGRLIDOKPISNL 1044
 Db 961 WTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNRLDGSKLYINGHLIDOKSILNL 1020
 Qy 1045 GNIHASNIMFKLDCGRDTHRYIWKYFNLPDKELNEKEIKDLYDNOSNGILKDFWGDY 1104
 Db 1021 GNIHVDNILFKIYNCST-RYIGIRVFNIFDKELDETEIOTLISNBPNTNLIKDFWGN 1079
 Qy 1105 LOYDKPYMLMLYDPNKYD-----VNVGIRGYMYLKGPRGSVWNTNIYLNSSLYR 1156
 Db 1080 LLYDKGYLLNVLKPNFIDRRDKDSTLSINNI-----RSTILLANRLYS 1123
 Qy 1157 GTFKFIKKY-ASGKNKINVRNDRVYINVVKNEYRLATNASQAGVEKILSALIEPDVG 1215
 Db 1124 GIKVKIORVNDSTNDRFRKNDQVYIYISNSSSYSLYADTNTTDKEKTIKS---SSSG 1180
 Qy 1216 N-LSQVVMKSKDQGITNCKMQLDNNGNDIGFIGHQPNNTAKLVASWYNRQIERS 1274
 Db 1181 NRFQVVMNS-----VGNCTMFKNNNGNIGLGF-----KADTVASTWYTHMRDH 1231
 Qy 1275 SRTLGSWEFIPVDGNGER 1294
 Db 1232 TNSNGCFWNFISEEHGQEK 1251

RESULT 8

BAB86845
 ID BAB86845 PRELIMINARY; PRT; 1252 AA.
 AC BAB86845
 DT 02-MAR-2004 (TREMBLrel. 27, Created)
 DT 02-MAR-2004 (TREMBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TREMBLrel. 27, Last annotation update)
 DE Botulinum neurotoxin type E.
 GN BONT/E.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=35396;
 RA Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
 RA Nakamura S., Katasawa T., Kozaki S.;
 RT "Sequence of the botulinum neurotoxin type E";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB082519; BAB86845.1; -;
 KW Neurotoxin.
 SQ SEQUENCE 1252 AA; 143637 MW; 76401D4D2E95D7A2 CRC64;

Query Match 34.3%; Score 2339; DB 2; Length 1252;
 Best Local Similarity 39.9%; Pred. No. 5.6e-96;
 Matches 535; Conservative 249; Mismatches 421; Indels 136; Gaps 43;

Qy 1 MPFVKQFNKYKDPVNGVDIAYIKIPNAGQMPQKAFIKHNKIWIPIRDTF-TNPEEGDL 59
 Db 1 MPKIN-SFNNDPVDNRTILYIK--PGCQEFYKSFNIMKNIWIPIERNVIGTTPQ--DF 55
 Qy 60 NPPEAKQVPVSYDSTYLSTDEKNYKLVGVTKLPERIYSTDLGRMLLTSIVRGIPFWG 119
 Db 56 HPPTSLKNGDSYYPNQLSQDEKDRFLKIVTKIFNRINNNLSGILLESKANPYLG 115
 Qy 120 G-STIDFELXVIDTNCINVIQPGSVRSSEELNLVLIQGSADIIQFECKSPGHEVLN--L 175
 Db 116 NDNTPDQFHIGASAVEIKFSGSQDILLPNVLIIMGAEPLD--FETNSSNISLRNNMP 173
 Qy 176 TRNGYGTQVIRFSPDFTFGFESLEVDTPNLLGAGKATDPVATLAHELIYAGHRLY-- 233
 Db 174 SNHGFGSIAIVTSPESYFRFNDN-----SMNEFIQDPALTLMEHLIHLGLYGA 224
 Qy 234 -GIAINPNRVKVN--TNAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNK 289
 Db 225 KGITTKYTIQKQNPILTN---IRGTNI--EEFLTFGGTDLNIIITSAQSNDIYTNLLAD 278
 Qy 290 FKDIASLANKAKSIVGTGTASIQYMKNVFKKYLISSEDTSGKFSVDKLFKDKLYKMLTEY 349
 Db 279 YKTIASKLSKVQV---SNPLNPKYQVFEAKYGLDKDASGIYSVINKNFNDFPKKLYS-F 334
 Qy 350 TEONFVFFKVLNKRKTYLNFDFKAVFKI-NIVPKVNYTYIDGFNLRNTNLAANFNGQNT 408
 Db 335 TEPDLATKQVKCQRTVIGQYK-YFKLSNLLNDSIYNISEGNI--NNLKVNFGQANL 391
 Qy 409 NNMNFTKLKNTG-----LFEFYK-LLCVRGITTSKSLDKGYNKALNDLCIKVNWMD 461
 Db 392 NPRIITPI---TGRGLVKKIIRFCNKIVSVKGIRKS-----ICIEINNGE 433
 Qy 462 LFFSPSEDNFTND-LNKGERI-----TSDTNIIEAENISLDLQOYVLTNFDNEPENIS 516
 Db 434 LFFVASENSYNDNINTPKEIDDTVTNNYE-----NDLQVILNFSEAP-GLS 484
 Qy 517 IENLSSDIIGOLELMPNIERPENG-----KKYELDKYTMFHYLRAQEFHGKSRALTNSV 572
 Db 485 DEKLNLTQND-AVLPKYD--SNGTSDIEQHDVNLNVPFFYLDQAKVPEGENNVNLTSSI 541
 Qy 573 NEALLNPSRVYTFSSDYVKKVKNKATEAAMFLGWVQLVDFDTDETSEVSTDKIADITI 632
 Db 542 DTALLEQPKIYTFSSSEFINNVKPVQAALFVSWIQVVLDFTEANQKSTVDKIADISI 601
 Qy 633 IIPYIGPALNIGNMLYKDDPFGALIFSGAVILLFPIEIPAIPLVGTALVSYIA---NKV 689
 Db 602 WVPYIGLALNIGNEAQKGNFKDALELGAGILLFVPELLIPTILVTFIKSFLGSSDNKN 661
 Qy 690 LTVOTIDNALS KRNEKWDVYKYIVTNWLAKVNTQIDLRKMKKEALENOAEATKALINY 749
 Db 662 KVIRAINNALKERDEKKEVYFVSNWMTKINTQFNKRKEQMYQALQVNAKLTIEF 721
 Qy 750 QYNQYTEEEKNI--NFNIDDLSSKLNESINKAMINIKFNQCSVSYLNMSPYGVYK 807
 Db 722 KNSYTTLEEKELKNNYDIEQIENELNQKVSIAAMNIDRFLTESSISYLMKLINEVKINK 781
 Qy 808 LEDFDASLKDALLKYIYDNRGLTIGQVDR-LKDKVNNTLSTDIPFOLSKYVDNQRLISTF 866
 Db 782 LREYDENVKYLLNYIIQH-GSILGESQBELNSVTDLTNNISIPFKLSSYTDKILISYF 840

[illegible]

Qy	176	TRNGYGTQVIRSPDPDTFGFBSLESDVTNPLIGAGKATDPDPAVTLAHELIYAGHRLY--	233
Db	177	SNHGFGSIAIVTSPSPFRKDN-----SMNEFIQDPALTMHLEIHTSLHGYGA	227
Qy	234	-GIAINPNRVKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEPRLYYNKK	289
Db	228	KGITTKYITQKONPLITN-----IRGTNI--BEFLTFGGTDLNIIITSAQSDNIYNLLAD	281
Qy	290	FKDIASLTNKAISIVGTFTASLQYMKVNVFKBKYLILSEDTSGKFSVDLKFDPKLYKMLTEIY	349
Db	282	YKKIASKLSKVQV---SNPLNPNPKDVFVEAKYGLDKDASGIYSVINKFNDFPKKLYS-F	337
Qy	350	TEDNFVKFKVLNRKTYLNFDPKAVEKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQTEI	408
Db	338	TEFDLATKFOVKCRQYIGQYK-YFKLSNLNDISIYNISEGYI--NNLKVNFRGQANL	394
Qy	409	NNNNFTKLNKFTG-----LPFEYK-LLCVRGIIITSKSLDKGYNKALNDLCIKVNNWD	461
Db	395	NPRIITPI---TGRGLVKKIIRFCNKIVSVKGIKRS-----ICIEINNGE	436
Qy	462	LFPSPESDNFTND-LNKGEI-----TSDTNIEAAEENISLDLLOQYYLTTFNDEPENIS	516
Db	437	LFFVASSENSYNDNIINTPKEDITDVTYSNNYE-----NDLQVTLNFNSESAP-GLS	487
Qy	517	IENLSSDIIGOLELMPNIERFPNG---KKVELDKYTMFYHLRAQEPHEGHSKRIALTNSV	572
Db	488	DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVDELNVFFYLDAQKVPBENNVNLTSSI	544
Qy	573	NEALLNPSRVYTFSSDYVKVKVNKATEAAMPLGWVEQLVYDFTDETSEVSTTKIADITI	632
Db	545	D'TALLEOPKIYTFSSSEFINNVKNPQOALFVGWIIQQVLVDFDTEANQKSTVDKIADISI	604
Qy	633	IIPYIGPALNIGMLYKDDPVGALIPSGAVILAEFPEIAIPVLGTFALVSYTA--NKV	689
Db	605	VVPYIGALNIGNEAQGNFKDLELUGALLEGELFEPPELLIPTILVETIISKFLGSSONKN	664
Qy	690	LTVQTDIINALSKENKWEYKIVTNWLAKVNTQDILRKQKKEALENOAEATKALINY	749
Db	665	KVIAKAINNALKERDEKWKVEYSVFVSNWMTKINTQFNKRKEQMVQALQNVNALKAIIES	724
Qy	750	OYNQYTEEEKNI--NFNIDDLSSKLNESINKAMININKELNQCYSVYLNMSMIPYGVKR	807
Db	725	KYNSYTLSEKXELNTKYDIEQIELNELNQKYSIAWNIIDRELTRESSIYMKLNEVKINK	784
Qy	808	LEDFDASLAKALLKYIDNRGTLIGQVDR-LKDKVANTLSTDIPQLSKYVDNQRLISTF	866
Db	785	LREYDENVKTYLLDYII-KHGSILGESQOELNSWVIDTLNNSIPFKLSSYTDCKILLSYF	843
Qy	867	TEYIKNIINTSILNLAYESNHLIDLRYASKINIGSKVNPDPDIDKXQIQULFNLESSKIEV	926
Db	844	NKFFPKRIKSSSVLNMRYKNDKYVDTSYDSDNININGDVYKYPTNKQFGIYNDKLSEVNI	903
Qy	927	ILKNAIVYASWYENFSTSWIRIPKYNSTI-SLNNEYTIINCH-ENNSGKWSVLYNTEGII	984
Db	904	SQNDYIITNDYKNKFSISFWVRIPNDVKLVNVNNEYTIINCRDRNNSGKWSVLNHNWEII	963
Qy	985	WTLQDTQEIQRVYFKYSQMINISDYINRWIFVTITNRLNNSKIYINGRLIDQKPISNL	1044
Db	964	WTLQDNGSINGQKLAIFYNGANGISDYINKWIFVTIITNDRLGDSKLYINGNLIDKKSILNL	1023
Qy	1045	GNTHASNNIMPKLDGCDGTHRYIWIWKYFNLFDEKLEKEKEIKLYDNCOSNGILKDFWGDY	1104
Db	1024	GNTHVSDNILFKIYVNCYSYT-RYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFPGNY	1082
Qy	1105	LOYDKPYMLNLYDPNKYVD-----VNVVGIRGYMWYLGKPRGSVMTTNIYNLSSLYR	1156
Db	1083	LLYDKEYYLLNVLPKNPNFINRRTDSTLSINNI-----RSTILLANRLYS	1126
Qy	1157	GTFKFIILKCY-ASGNKONIVRNRDRVYINVVVKNEYRLATNASQAGVEKILSALIEIPDVG	1215
Db	1127	GIVKIORVNNSTNDMLVRKNDQVYINFA-SKTHLLPYADYATTNK-EKTIKILSSG	1184

RT gene.";
 RL Eur. J. Biochem. 204:657-667(1992).
 RN [3]
 RP SEQUENCE OF 1-251 FROM N.A.
 RX MEDLINE=90264400; PubMed=2160960;
 RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
 RT "The complete sequence of botulinum neurotoxin type A and comparison
 with other clostridial neurotoxins.";
 RL J. Biol. Chem. 265:9153-9158(1990).
 RN [4]
 RP SEQUENCE OF 1-13.
 RX MEDLINE=85197963; PubMed=3888113;
 RA Schmidt J.J., Sathiamoorthy V., Dasgupta B.R.;
 RT "Partial amino acid sequences of botulinum neurotoxins types B and
 E.";
 RL Arch. Biochem. Biophys. 238:544-548(1985).
 RN [5]
 RP SEQUENCE OF 419-426.
 RX MEDLINE=90344918; PubMed=2116911;
 RA Gimenez J.A., Dasgupta B.R.;
 RT "Botulinum neurotoxin type E fragmented with endoprotease Lys-C
 reveals the site trypsin nicks and homology with tetanus neurotoxin.";
 RL Biochimie 72:213-217(1990).
 RN [6]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94063091; PubMed=8243676;
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
 RA Benfenati F., Wilson M.C., Montecucco C.;
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
 COOH-terminal peptide bonds.";
 RL FEBS Lett. 335:99-103(1993).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94124495; PubMed=8294407;
 RA Binz T., Blas J., Yamazaki S., Baumeister A., Link E., Suedhof T.C.,
 RA Jahn R., Niemann H.;
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
 RL J. Biol. Chem. 269:1617-1620(1994).
 CC -I- FUNCTION: Botulinum toxin acts by inhibiting neurotransmitter
 release. It binds to peripheral neuronal synapses, is internalized
 and moves by retrograde transport up the axon into the spinal cord
 where it can move between postsynaptic and presynaptic neurons. It
 inhibits neurotransmitter release by acting as a zinc
 endopeptidase that catalyzes the hydrolysis of the 180-Arg-Ile-
 181 bond in SNAP-25.
 CC -I- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
 neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
 detected action on small molecule substrates.
 CC -I- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -I- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
 heavy chain (H). The light chain has the pharmacological activity,
 while the N- and C-terminal of the heavy chain mediate channel
 formation and toxin binding, respectively.
 CC -I- SUBCELLULAR LOCATION: Secreted.
 CC -I- MISCELLANEOUS: There are seven antigenically distinct forms of
 botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
 CC -I- SIMILARITY: Belongs to peptidase family M27.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X62089; CAA43999.1; -;
 DR EMBL; X62683; CAA44558.1; -;
 DR PIR; S08575; S08575.
 DR PIR; S21178; S21178.
 DR HSP; Q45894; 1EH.
 DR MEROPS; M27.002; -;
 DR InterPro; IPR008985; ConA_like_lec_gl.

DR InterPro; IPR011065; Kunitz like.
 DR InterPro; IPR000395; Peptidase M27.
 DR InterPro; IPR006025; Pept M_Zn_BS.
 DR Pfam; PF01742; Peptidase_M27; 1.
 DR PRINTS; PR00760; BONTOKILYSIN.
 DR PRODOM; PD001963; Bontokilysin; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Direct protein sequencing; Hydrolase; Metalloprotease; Neurotoxin;
 KW Transmembrane; Zinc.
 FT INIT_MET 0
 FT CHAIN 1 421 Botulinum neurotoxin E light-chain.
 FT CHAIN 422 1250 Botulinum neurotoxin E heavy-chain.
 FT METAL 211 211 Zinc (catalytic) (By similarity).
 FT ACT_SITE 212 212 By similarity.
 FT METAL 215 215 Zinc (catalytic) (By similarity).
 FT DISULFID 411 425 Interchain (Probable).
 FT CONFLICT 176 176 R -> G (in Ref. 2).
 FT CONFLICT 197 197 C -> S (in Ref. 2 and 3).
 FT CONFLICT 339 339 R -> A (in Ref. 2).
 FT CONFLICT 772 772 I -> L (in Ref. 2).
 FT CONFLICT 962 963 FE -> LQ (in Ref. 2).
 FT CONFLICT 966 966 R -> A (in Ref. 2).
 FT CONFLICT 1194 1194 N -> NN (in Ref. 2).
 SQ SEQUENCE 1250 AA; 143712 MW; D9FCE26DDA041EB4 CRC64;
 Query Match 33.9%; Score 2309.5; DB 1; Length 1250;
 Best Local Similarity 39.6%; Pred. No. 1.2e-94;
 Matches 530; Conservative 250; Mismatches 423; Indels 137; Gaps 43;
 QY 2 PFVNKQPNYKDPNGVNDIAYIKIPNAGQMPQVAFKXIHNNKIWIPIERDTF-TNPEEGDLN 60
 DB 1 PKIN-SFNNDPVNDRTILYIK-PGCGCFYKSFNIMKNIWIPIERNVIGTTPQ--DFH 55
 QY 61 PPEAKQVPVSYVDSTYSLTDNEKDNVYKGVTKLPERIYSTDGRMLTTSIVRGIPWGG 120
 DB 56 PPTSLKNGDSYDDPNYLOSDEEKDRFLKTVTKFNRNNLNKSGGILLEELSKANPYLGN 115
 QY 121 -STIDYELKVIDTNCINVIQDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LT 176
 DB 116 DNTPDQNFHGDASAVEIKESNGSQDILLPNVIMGAEPDL--FETNSSNISLRNNWPS 173
 QY 177 RNYGSGTQYIRSPDFTFGPEESLEVDNTPNLLGAGKAPDPAVTLAHELIYAGHRLY--- 233
 DB 174 NHRFGSIAITVTSPEYSFRENDNC-----MNEFIQDPALTLMHELIHSLHLYGAK 224
 QY 234 GIAPNPRVFKVN--TNAYENSGLEVSFEELRTFGGHDKEFIDSLQENEFRLYYNKF 290
 DB 225 GITTKYITOKQPLITN-----IRGTNI--EEFLTFGGTDLNITSAQSNDIYTNLLADY 278
 QY 291 KDIASLTINKAKSIVGTTASIQYMKNVFKKYLSEDTSGKFSVDKLFKLYKMLTEIYT 350
 DB 279 KTIASKLSKVQV---SNPLNPKYDFEAKYGLDKDASGIYSVNINKNFDFKLYS-FT 334
 QY 351 EDNVPKFFKVLNRTKTYLNFDKAVFKI-NIVPKVNYITYDGFNLRNLTNLAANFNGQTEIN 409
 DB 335 EFDLRTKQVQKQRTYIGQYK-YFKLSNLINDSIYINISEGYNI--NNLKVNFRGQANLN 391
 QY 410 NMFTKLNKFTG-----LFEFYK-LLCYRGIIITSKTSKLDKGNKALNDLCIKVNNWDL 462
 DB 392 PRIITPI---TGRLVKKIRFCNKIVSVKGIKRS-----ICIEINNGEL 433
 QY 463 FFSFSDNFTND-LNKGEET---TSDTNI EAABENISLDLIQQYYLITFNDPENENISI 517
 DB 434 FFVASENSYNDNDINTPKETDDTTSNNYE-----NDLDQVILNFSSESAP-GLSD 484
 QY 518 ENLSSDIIGOLELMPNIERPFG-----KKYELDKYTMFHYLRAQEFHGKSRALTNSVN 573
 DB 485 EKLNLATQND-AIYPKYD--SNGTSDIEQHDVNLNLFVFFYDLAQKRVPEGENNLTSSID 541
 QY 574 EALINPFRVYTFSSDYVKKVNKATAAAMFLGWVEQLVYDFTDETSEVSTDTKIADITII 633
 DB 542 TALLEQPKIYTFSSSEFINNVNKPQALFVSIQQVQLVDFVTTEANQKSTVDKIADISIV 601


```
FT DISULFID 429 445 Interchain (Probable).
SQ SEQUENCE 1274 AA, 146709 MW, 5899756A7438B921 CRC64;

Query Match 33.7%; Score 2296.5; DB 1; Length 1274;
Best Local Similarity 39.4%; Pred. No. 4.5e-94;
Matches 523; Conservative 253; Mismatches 465; Indels 85; Gaps 39;

Qy 1 MPFVNKQNYKDPVNGVDIAIKIPNAGOMQ-P-KAFKIHKNKIWIPIERDTF-TNPBEGD 58
Db 1 MPVAINSPYNDPVNDITLWQIPYEKSKYKAFELMENWIIIPERTIGNP--SD 58

Qy 59 LNPPEAKQVSVYSDYSTLSTNEKNYKLVKGVTKLPERIYSTDGLRMLLTSIVRGIPFW 118
Db 59 FDPASLKNQSSAYDPNLTDAEKORYLTKTKFKRINSNPAGKVLQEISVAKPYL 118

Qy 119 GG--STDTLKVLTNCINVIQDGSYRSELANVIIGPSADIIQPEC-----KSFGEV 172
Db 119 GNDHTPIDFSPVTRTTSVNTKLSNTVESSMLLLVLGAGDIPFESCCYVPRKLDIPDV 178

Qy 173 LNLTRN-GYGSTQVIRSPDFTFGFEESLEVDTNPLLCAGKGFADPAVTLAHELIYAGHR 231
Db 179 VYDPSNYGFGSINIVTSPVEYTFN---DISGCHNSSTESFIADPAISLAHELIHALHG 235

Qy 232 LYGI-AINPNRVKNTNAYEMSGLEVSFEELTFGGHDAKFIDSIOENEFRLYYNKF 290
Db 236 LYGARGVTEETIEVK-QAPLMIAEKPIRLEFLTFGQDILNIIITSAMKEKIYNNLLANY 294

Qy 291 KDIASLTNKAISVGTASLOY---MKGVPEKYLSEDSGKFSVDKLGKLYKWLIT 346
Db 295 EKIAITRUSEVNS-----APPEYDINEYKDYQWYGLDKDNADGYSVTYNNENKFBIKKLY 349

Qy 347 EITYEDNPFVKFVLRNRYLNFDAKVF-KI-NIVPKVNYTYDGFNLNRYNTLNAFNQG 404
Db 350 S-FIESDLANKFKVCRNTY--FIKYEFLKVPNLLDDDIYTVSGFNI--GNLAVNRGQ 404

Qy 405 NTEINNMTFLKNFTGLFEPYKLLCVRGITTSKTSKLDKGNKALNDLCHIKVNNWDLFF 464
Db 405 SIKLNPKIIDIPD-KGLIVERIKVFC-KSVIPRKG-----TKAPPLRCIRVNNSELF 455

Qy 465 SPSEDNFT-NDLNGEERTSTDTNEAEENISLDLIQOYLYLTFNDEPENISENISSD 523
Db 456 VASSESYNENDINTPKEDIDTNNNNYRN-NLD---EVILDYNSQTIPO-ISNRTLNT- 509

Qy 524 IIGOLELMPNIERFPNG----KTYELDKYTMFHYLRAQEPFHGKSRIALTNSVNEALNP 579
Db 510 LVQDNSVYPRVD--SNGTSEIEYDVVDVNFVFFYLHAQKVEGETNISLSTSSIDTALLEE 567

Qy 580 SRVTFSSDYVKKVKAATEAMFLGWVEQLVYDFTDETSEVSTDKIADITIIPYGP 639
Db 568 SK-DIFFSESPIETINKPVNAALFDWISKVIRDFTTEATQKSTVDKIADISLIVPVYGL 626

Qy 640 ALNIGNMLYKDDFVICALIFSGAVILLPEIPATIPVLGTALVSVI---ANKVLTVQTI 696
Db 627 ALNIIIEBAKGNFEAFELLVGILLPEVPBELTIPVILVFTIKSYIDSYENKKNKAIKAIN 686

Qy 697 NALSKRNEKWDVYKVIYVNNLAKVNTQIDILIRKMKKEALENQAEYKAIINYQNOYTE 756
Db 687 NSLIEREAKWKEIYSWISVNWLTINTQFNCKRQSMYQALQNOVDAIKTALEYKNNYTS 746

Qy 757 BEKNNI--NFNIDLSKLNESINKAMINKNFLNQCSVSVLANSMTPIYGVKRLLEDFA 814
Db 747 DEKNRLESEYNNIEBELNKKVSLAMKNIERFERFTESSISYLMKLINEAKVGLKKYKDNH 806

Qy 815 LKDALIKYIDNRGLIGQVDRDKVNNLTSTDIIPQLSKYVDNORLLSTFTEVIKNI 874
Db 807 VKSDLLNVLDRSILGEQTNELSDLTSTLSSNIPFELSSTYNDKILIIYFNRLYKXIK 866

Qy 875 NTSILNRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIOFLNLESSKIEVLKNAIV 934
Db 867 DSSILDWRYENKFKIDISGYGNSISINGNVYIYSTNRNFGIYNSRLSEVNIAQNNDIY 926

Qy 935 NSMYENSTSWAIRPKPFGNISLNNEYTIINCM-ENNSGKVSILNY---GEIITWLODT 990
Db 935 NSMYENSTSWAIRPKPFGNISLNNEYTIINCM-ENNSGKVSILNY---GEIITWLODT 990
```

RESULT 15

Q8GR96

```
ID Q8GR96 PRELIMINARY; PRT; 1291 AA.
AC Q8GR96;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Neurotoxin.
GN Name=bontib;
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RA Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA Nakamura S., Mukamoto M., Kozaki S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB084152; BAC22064.1; --
DR HSP; P10844; IEPW.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR011591; Botulinum.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR011065; Kunitz-like.
DR InterPro; IPR000395; Peptidase M27.
DR InterPro; IPR006025; Pept M Zn BS.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR ProDom; PD001963; Botulinum; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR Neurotoxin.
SQ SEQUENCE 1291 AA; 150574 MW; 0227CAEF4F58504D CRC64;
```

Query Match 33.6%; Score 2290; DB 2; Length 1291;

Best Local Similarity 39.1%; Pred. No. 8.9e-94;

Matches 523; Conservative 232; Mismatches 490; Indels 94; Gaps 25;

Qy 1 MPFVNKQNYKDPVNGVDIAIKIPNAGOMQ-P-KAFKIHKNKIWIPIERDTF-TNPBEGD 58

Db 1 MPVTINNFYNDPIDNNIIMPEPPFARGTGRYKAFKIDRIWIIPERTYTFYKPBDFN 60

Qy 59 LNPPEAKQVSVYSDYSTLSTNEKNYKLVKGVTKLPERIYSTDGLRMLLTSIVRGIPFW 118

Db 61 KSSGIFNRDV-CRYDDPYLNTNDKNIFLQTKNIFLNRIKSKPLGKLEMIINGIPYL 119

This Page Blank (uspto)